Second-Order Least Squares Estimation in Dynamic Regression Models

by

Mustafa Kamal AbdelAziz Salamh

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of

Doctor of Philosophy

Department of Statistics
University of Manitoba
Winnipeg
Canada

Abstract

In this dissertation we proposed two generalizations of the Second-Order Least Squares (SLS) approach in two popular dynamic econometrics models. The first one is the regression model with time varying nonlinear mean function and autoregressive conditionally heteroskedastic (ARCH) disturbances. The second one is a linear dynamic panel data model. We used a semiparametric framework in both models where the SLS approach is based only on the first two conditional moments of response variable given the explanatory variables. There is no need to specify the distribution of the error components in both models.

For the ARCH model under the assumption of strong-mixing process with finite moments of some order, we established the strong consistency and asymptotic normality of the SLS estimator. It is shown that the optimal SLS estimator, which makes use of the additional information inherent in the conditional skewness and kurtosis of the process, is superior to the commonly used quasi-MLE, and the efficiency gain is significant when the underlying distribution is asymmetric. Moreover, our large scale simulation studies showed that the optimal SLSE behaves better than the corresponding estimating function estimator in finite sample situation. The practical usefulness of the optimal SLSE was tested by an empirical example on the U.K. Inflation.

For the linear dynamic panel data model, we showed that the SLS estimator is consistent and asymptotically normal for large N and finite T under fairly general regularity conditions. Moreover, we showed that the optimal SLS estimator reaches a semiparametric efficiency bound. A specification test was developed for the first time to be used whenever the SLS is applied to real data. Our Monte Carlo simulations showed that the optimal

SLS estimator performs satisfactorily in finite sample situations compared to the first-differenced GMM and the random effects pseudo ML estimators. The results apply under stationary/nonstationary process and wih/out exogenous regressors. The performance of the optimal SLS is robust under near-unit root case. Finally, the practical usefulness of the optimal SLSE was examined by an empirical study on the U.S. airfares.

Acknowledgement

I would like to offer my gratitude to Prof. Liqun Wang, my advisor, who accepted to supervise me, and has guided me throughout my Ph.D research with patience and knowledge. I also thank him for supporting me financially through his Natural Sciences and Engineering Research Council of Canada (NSERC) grant. It was very nice to work with him over the past five years.

I am gratefully acknowledging the financial support provided through the University of Manitoba Graduate Fellowship (UMGF), Manitoba Graduate Scholarships (MGS).

I thank the members of my advisory committee, who provided me with their suggestions to improve this research: Prof. James Fu, Prof. Umut Oguzoglu at University of Manitoba.

Last but not least, I thank the Department of Statistics at University of Manitoba for hosting me over the past five years.

Dedication

This thesis is dedicated to my beloved parents, wife and daughters

Contents

List of Tables

List of Figures

Overview

In these few pages we try to give an overview about the method of second-order least squares (SLS) as a moment-based method of estimation; how it was introduced, developed, and generalized to different areas of applications. Then we give some motivations to this current research, and conclude by describing the structure of this dissertation.

The SLS estimator was firstly introduced in the literature by Wang (2003), who used to refer to it at early stage as a minimum distance moment-based estimator. This is because it minimizes simultaneously the distances of the response variable and the squared response variable to the first and second conditional moments of the response variable given the observed covariates. The motivation was that in many situations, parameters in nonlinear regression models with Berkson-type measurement errors in one predictor can be identified and, therefore, consistently estimated using the first two conditional moments of the response variable given an observed (proxy) variable. The consistency and asymptotic normality of the SLS was shown under the assumptions that the measurement errors are iid normal and the random disturbances in the regression equation are iid with unspecified distribution

The idea of SLS estimation was extended in (Wang, 2004) to cover the case of several predictors with Berkson-type measurement errors, wherein the normality assumption on the measurement errors was relaxed by assuming a general parametric distribution, which is not necessarily normal. A refinement to the SLS was suggested by including a weight matrix in the objective function to attain more asymptotic efficiency. Moreover, to overcome the possible computational difficulty of minimizing an objective function which involves multiple integrals, a simulation-based estimator was constructed. Consistency and asymptotic normality for both estimators are derived under fairly general regularity conditions.

The name of SLS started to appear in the literature by 2007, where the estimator was used in a longitudinal data framework after it had been used in a cross sectional framework in the previous two papers. As a kind of multivariate generalization, Wang (2007) considered the SLS in a unified estimation framework which covers the nonlinear mixed effects models and the general nonlinear regression models with Berkson measurement errors in the covariates. The random effects (or measurement errors) had a general parametric distribution, whereas the unobserved predictor variables and disturbance terms had nonparametric distributions. To make use of the longitudinal structure, the SLS was based on the first conditional moments and all possible second order conditional moments of the response variables given the observed covariates. The author also handled a general case where the closed forms of the first two conditional moments are difficult or impossible to obtain, he proposed a simulation-based estimator and showed that both estimators are consistent and asymptotically normally distributed under fairly general regularity conditions. Limited Monte Carlo simulation studies were conducted, and it was shown that the SLS estimator with optimal weight matrix; optimal SLS, performs fairly satisfactorily for relatively small sample sizes and slightly better than the

Generalized Estimating Equation (GEE) estimator.

Wang and Leblanc (2008) investigated the SLS for a general nonlinear regression models with cross sectional data, where the random errors have homoscedastic variance and unknown distribution. They showed that the optimal SLS estimator is asymptotically more efficient than the ordinary least squares estimator if the third moment of the random error is nonzero, and both estimators have the same asymptotic covariance matrix if the error distribution is symmetric. Simulation studies showed that the variance reduction of the optimal SLS can be as high as 50% for sample sizes lower than 100.

Since 2008 there have been different attempts to apply the method of SLS and study its properties in different frameworks. We refer to most of them herein.
The first attempt was by Abarin and Wang (2009) who applied the SLS approach to the Tobit (censored regression) model where the error term has a general parametric distribution (not necessarily normal). They showed the strong consistency and asymptotic normality of the estimator and its simulation-based version under fairly general regularity conditions. They also studied the finite sample behavior of the SLS estimator with either identity or optimal weight matrices and compared the two estimators with the ML estimator over a range of error distributions and censoring degrees. They concluded that both of the ML and optimal SLS have close performance for the case of correctly specified model, however the SLS with identity weight matrix is more robust in case of models with misspecified error distribution.

The second attempt was by Abarin and Wang (2012) who used the the SLS approach in the estimation of generalized linear models with classical errors-in-variables in the predictor variables. They used the instrumental variable equation to build up the estimating moments. The error in the instrumental equation has parametric distribution that is not necessarily normal, while the distributions of the unobserved covariates, and the measurement errors are nonparametric. They also proposed simulation-based estimators for the situation where the closed forms of the moments are not available. They showed that the proposed estimators are strongly consistent and asymptotically normally distributed under some regularity conditions. Simulation studies showed that the estimators perform satisfactorily in some finite sample situations.

The third attempt was by Li and Wang (2012a) who used the SLS approach in the framework of generalized linear mixed models. They proposed a strongly $\sqrt{n}$-consistent simulation-based estimator which is based on the first two marginal moments of the response variables, and it allows the random effects to follow a flexible distribution (not necessarily normal). They showed the robustness of their estimator against data outliers. They also suggested a kind of Jackknife technique to build up the optimal weight matrix. This is in order to reduce the bias which results from approximating the optimal weight matrix. According to their simulations, the proposed estimator has desirable finite sample properties in comparison with the likelihood-based methods.

Last application but not least, Li and Wang (2012b) extended the work of Abarin and Wang (2012) to the generalized linear mixed models with classical measurement error in the covariates. They constructed a simulation-based estimator by combining the method of instrumental variables and the method of moments. Their proposed approach does not require either the parametric assumptions for the distributions of the unobserved covariates or the normality assumption for the random effects. The strong consistency and asymptotic normality of the estimators are obtained under mild regularity conditions. They run three simulations and concluded that their proposed estimator is almost unbiased compared to the sever biased found in the naive ML estimator that ignores the measurement error.

The semiparametric efficiency of the SLS was investigated for the first time by Kim and Ma (2012) who showed that Wang and Leblanc (2008) estimator reaches the optimal efficiency bound in the sense of Bickel et al. (1993). They used the geometric approach to identify the optimal semiparametric efficient (SE) estimator in the nonlinear regression model with homoscedastic error variance. Then they derived the estimation variance of the SE estimator, which appears to be the same as the asymptotic variance of the optimal SLS. This demonstrated that the optimal SLS estimator is semiparametrically efficient under this setup.

From all what have been mentioned it is evident that most of the studies on the SLS were done in a cross sectional framework, and even those studies which treated the case of longitudinal data models were investigating the asymptotic properties with regard to the cross sectional dimension. More crucially, all of these studies assumed strict exogeneity of the explanatory variables. Therefore, all of the asymptotic and finite sample results obtained so far don't apply either in case of dealing with general stochastic process or having endogenous variables in the regression equation. This raises up two sets of questions to be answered in this study.

First, how can the SLS be defined in a general stochastic (not iid) framework?, what are the sufficient regularity conditions to derive the asymptotic properties such as consistency and asymptotic normality of this estimator?, and what are the large sample and small sample merits of this estimator over the other commonly used estimating approaches such as the Quasi Maximum Likelihood (QML) and the Estimating Function (EF)?

Second, how does the SLS approach work in dynamic panel data regression models?, how to make use of the SLS estimation framework to deal with the built-in endogeneity resulting from regression on the lags of the response variable?, and how much gain of efficiency of SLS in large and finite samples compared to other commonly used estimating approaches such as the Generalized Method of Moments (GMM), Conditional Generalize Lest Squares (CGLS), and other pseudo likelihood based methods.

This dissertation comes in two main chapters in addition to the conclusion. In the first chapter we study the SLS approach for a general nonlinear regression model with time variant dynamic mean function and autoregressive conditionally heteroskedastic (ARCH) disturbances. We provide answers to the first group of questions. In the second chapter we provide answers to the second group of questions by investigating the SLS approach in linear dynamic panel data model. All of the mathematical proofs and supplementary definitions and lemmas are found in Appendix A, while a sample of R programming code is included in Appendix B.

For the sake of clarity in notation, the little bold letters are preserved for vectors, while the capital bold letters are preserved for matrices. The dimensions of a matrix or vector are given only if it is relevant. Each notation is defined locally within each chapter.

Chapter 1

SLS Estimation in ARCH Non-Linear Regression Model

Since the seminal work of Engle (1982), the autoregressive conditional heteroscedasticity (ARCH) model and its various generalizations have been intensively studied and widely used to analyze time series data, especially in economics and finance. The ARCH regression models allow both the conditional means and variances of a process to jointly evolve over time. Engle (1982) showed that the relative efficiency of the maximum likelihood (ML) estimator compared to the ordinary least squares estimator can be infinite under the conditional normality of the disturbances. Latter, Weiss (1986) showed that violating the normality assumption does not affect the consistency of the Gaussian quasi-ML estimator for the ARCH(p) regression models. However, Engle and Gonzalez-Rivera (1991) found that in a GARCH(1,1) model the asymptotic variance of the QMLE can be six or two times larger than Cramer-Rao bound if the conditional distribution of the disturbances is highly skewed or it exhibits leptokurtosis respectively. They also concluded that there was enough empirical and theoretical evidence to reject the assumption of conditional normality in financial time series, and that it was worthwhile searching for estimators that could improve the QMLE. Moreover, they proposed a semiparametric estimator by using the data to approximate the true generating mechanism of the disturbance term and pointed out that their estimator did not capture the total potential gain of efficiency. This finding motivated Li and Turtle (2000) to examine the estimating functions (EF) approach. Compared to the asymptotic procedures, they combined linear and quadratic EFs optimally based on the information criterion of Godambe (1985). They found a significant gain of efficiency from the EF approach over the quasi-likelihood approach in the case of serious departures from normality. On the same line, Liang et al. (2011) emphasized the benefit of using the first four conditional moments of the observed process in deriving a quadratic EF which maximizes Godambe's information criterion.

So far most of the research in the literature focuses on the theory of QML for different generalizations of ARCH regression models. To the best of our knowledge, the paper of Li and Turtle (2000) is the only one which studied in some details the application of a moment-based approach in ARCH linear regression models. Moreover, almost all the developed theory in the literature is based on the assumption of having mean stationary data generating process, specifically ARMA process. This is not general enough to cover models with time varying nonlinear mean function. Such models are very useful in capturing the nonlinear dynamic behaviour in time series data without doing any transformations on the variables of interest, and hence they preserve any structural relationship from being altered or weakened throughout the analysis. As stated by Enders (2010) "Economic theory suggests that a number of important time series variable should exhibit nonlinear behaviour...such as wages and employment". Similarly, Franses and Van Dijk (2000) mentioned that "financial time series display typical nonlinear characteristics". This assures that nonlinear time series can be seen frequently in economics and finance, and there is a need to study in some details the ARCH nonlinear regression models for the sake of more generality and flexibility.

In this chapter we attempt to fill these theoretical gaps. In particular, we consider a model with ARCH(p) disturbances and a fairly general nonlinear mean function that is allowed to be time varying. For this model, we study the SLS approach which is based on the first two conditional moments of the process. An optimal estimator is obtained by using the additional information inherent in the conditional skewness and kurtosis of the process. The consistency and asymptotic normality of the SLS estimator are established under general mixing conditions. We demonstrate that SLS approach leads to efficiency gain over the commonly used QMLE and the gain is significant under asymmetric distributions. The practical usefulness of this gain of efficiency is emphasized by an empirical example. Our extensive simulation results show that the optimal SLS estimator has in most cases superior finite sample properties over the EF approach based on the same set of moments.

This chapter is organized as follows. In section 1.1, the model is introduced and the SLS estimator is defined. In section 1.2, the strong consistency and asymptotic normality of the SLS estimator are shown. In section 1.3, the optimal SLS estimator is derived and a feasible optimal SLS estimator is suggested. In section 1.4 we investigate the gain of efficiency of the optimal SLS estimator relative to the QMLE and highlight the differences between our approach and the EF approach. In section 1.5 we perform Monte Carlo analysis to examine the behavior of the SLS estimator in terms of its bias and root mean squared error (RMSE) in the cases of small and moderate sample sizes and for both skewed and symmetric distributions of the disturbances. Section 1.6 demonstrates the merits of the SLS approach using the U.K. price Inflation example of Engle (1982). A summary is given in section 1.7.

1.1 Model Specification and SLS Estimation

Let $\left\{(\boldsymbol{x}_t', y_t)'\right\}$ denote a sequence of random vectors defined on a complete probability space $(\Omega, \mathcal{F}, P)$ and let $\mathcal{F}_{t-1} = \mathcal{F}\left\{\boldsymbol{x}_i', y_{i-1},\, i \leq t\right\}$. For some non-negative integer τ, let $\boldsymbol{v}_t = \left(\boldsymbol{x}'_{t-\tau}, y_{t-\tau}, \ldots, \boldsymbol{x}'_{t-1}, y_{t-1}, \boldsymbol{x}'_t\right)'$. Assume that y_t can be represented as

$$y_t = f_t\left(\boldsymbol{v}_t, \boldsymbol{\theta}_0\right) + \epsilon_t, \quad t \in \mathbb{Z}, \tag{1.1}$$

where $f_t : \mathbb{R}^v \times \Theta \to \mathbb{R}^1$, are known functions measurable on $\mathbb{R}^v$ for each $\boldsymbol{\theta}$ in Θ (a subset of $\mathbb{R}^q$), and continuous on Θ uniformly in t a.s.-P. Let $\epsilon_t = \sigma_t \varepsilon_t$, such that

$$E\left(\varepsilon_t \left|\mathcal{F}_{t-1}\right.\right) = 0 \quad \text{a.s.-}P, \quad E\left(\varepsilon_t^2 \left|\mathcal{F}_{t-1}\right.\right) = 1 \quad \text{a.s.-}P, \quad \sigma_t^2 = \phi_{00} + \sum_{i=1}^{p} \phi_{0i}\epsilon_{t-i}^2, \tag{1.2}$$

where ϕ_{00}, $\phi_{0p} > 0$, and $\phi_{0i} \geq 0$ for $i = 1, 2, \ldots, p-1$. Model model (1.1) - (1.2) contains the ARCH linear regression model as a special case. Our main interest is to estimate $\boldsymbol{\gamma}_0 = (\boldsymbol{\theta}'_0, \boldsymbol{\phi}'_0)'$ in Γ (a compact subset of $\mathbb{R}^{q+p+1}$) based on a realization of $\left(\boldsymbol{x}'_{1-p-\tau}, y_{1-p-\tau}, \ldots, \boldsymbol{x}'_T, y_T\right)'$.

According to the above setup, we can use Theorem 2.11 of White (1996) to show the existence of functions $\hat{\boldsymbol{\gamma}}_T : \Omega \to \Gamma$ measurable–$\mathcal{F}$, $T = 1, 2, \ldots$ such that

$$\hat{\boldsymbol{\gamma}}_T = \underset{\boldsymbol{\gamma} \in \Gamma}{\operatorname{argmin}}\, Q_T\left(\boldsymbol{\gamma}\right) \quad \text{a.s.-}P, \tag{1.3}$$

where $Q_T\left(\boldsymbol{\gamma}\right) = T^{-1}\sum_{t=1}^{T} \boldsymbol{h}'_t(\boldsymbol{\gamma})\boldsymbol{W}_t\boldsymbol{h}_t(\boldsymbol{\gamma})$, given that $\boldsymbol{W}_t$ is a 2×2 matrix which is measurable with respect to $\mathcal{F}\left\{\boldsymbol{v}'_{t-p}, \ldots, \boldsymbol{v}'_t\right\}$ and non-negative definite a.s.-P, $\boldsymbol{h}'_t(\boldsymbol{\gamma}) = (\epsilon_t(\boldsymbol{\theta}),\, y_t^2 - f_t^2\left(\boldsymbol{v}_t, \boldsymbol{\theta}\right) - \sigma_t^2(\boldsymbol{\gamma}))$, $\epsilon_t(\boldsymbol{\theta}) = y_t - f_t\left(\boldsymbol{v}_t, \boldsymbol{\theta}\right)$, and $\sigma_t^2(\boldsymbol{\gamma}) = \phi_0 + \sum_{i=1}^{p} \phi_i \epsilon_{t-i}^2(\boldsymbol{\theta})$.

1.2 Asymptotic Properties of the SLS Estimator

In this section we establish the consistency and asymptotic normality of $\hat{\boldsymbol{\gamma}}_T$ under fairly general assumptions. Towards that, we adopt the following mixing condition in order to restrict the memory of the underlying process and guarantee a sort of asymptotic independence.

Assumption 1 *The process* $\left\{(\boldsymbol{x}_t', y_t)'\right\}$ *is strong mixing of size* $-a$*, for some* $a > 1$*. (See definition 6 in appendix A).*

This is a high-level operating assumption which allows for considerable dependence and heterogeneity in the underlying process. As mentioned by White and Domowitz (1984), this single assumption preserves the asymptotic independence of the observed process even under further transformations. The assumption can be justified on a case by case basis. For example, Lindner (2009) stated that if $\sum_{i=1}^{p} \phi_{0i} < 1$, and the driving noise sequence $\{\varepsilon_t\}$ is i.i.d. and absolutely continuous with Lebesgue density being strictly positive in a neighbourhood of zero and finite second moment, then $\{\epsilon_t\}$ is strong mixing with geometric rate. The geometric memory decay implies that a can be set to arbitrary large number. It is also shown that finite order Gaussian ARMA processes are strong mixing (Ibragimov and Linnik, 1971, pp. 312–313).

The consistency of $\hat{\boldsymbol{\gamma}}_T$ follows from the uniform convergence of $\{Q_T(\boldsymbol{\gamma})\}$ (on Γ) to a non-stochastic sequence $\left\{\bar{Q}_T(\boldsymbol{\gamma})\right\}$ which possess unique minimizers at $\boldsymbol{\gamma}_0$ for all T sufficiently large. To fulfil that, we impose assumptions 2, and 3 respectively.

Assumption 2 *Let* $\|.\|$ *be the Euclidean norm, then we have*

$$\sup_{t\in\mathbb{N}} E\left\{\|\boldsymbol{W}_t\|\left(1+\sum_{i=0}^{p}\epsilon_{t-i}^4+\sup_{\Theta} f_{t-i}^4\left(\cdot,\boldsymbol{\theta}\right)\right)\right\}^r < \infty,\ \textit{for some}\ r > \frac{a}{a-1}.$$

By using Hölder's inequality and c_r inequality, we can easily verify that the sequence $\{\boldsymbol{h}_t'(\boldsymbol{\gamma})\boldsymbol{W}_t\boldsymbol{h}_t(\boldsymbol{\gamma})\}$ is dominated by uniformly L_r-bounded variables[1]. Therefore, $\overline{Q}_T(\boldsymbol{\gamma}) = T^{-1}\sum_{t=1}^{T} E\{\boldsymbol{h}_t'(\boldsymbol{\gamma})\boldsymbol{W}_t\boldsymbol{h}_t(\boldsymbol{\gamma})\}$ is well defined and is continuous on Γ uniformly in T. Moreover, by the uniform law of large numbers in (White and Domowitz, 1984, Theorem 2.3) we have $\sup_{\boldsymbol{\gamma}\in\Gamma}\left|Q_T(\boldsymbol{\gamma}) - \overline{Q}_T(\boldsymbol{\gamma})\right| \xrightarrow{a.s.} 0$ as $T \to \infty$.

Assumption 3 *For any open neighbourhood $\mathcal{N} \subsetneq \Gamma$ of $\boldsymbol{\gamma}_0$, there exists $T_0(\mathcal{N})$ such that*

$$\inf_{T\geq T_0}\left(T^{-1}\sum_{t=1}^{T}\min_{\boldsymbol{\gamma}\in\mathcal{N}^c\cap\Gamma} E\left\{(\boldsymbol{h}_t(\boldsymbol{\gamma}) - \boldsymbol{h}_t(\boldsymbol{\gamma}_0))'\,\boldsymbol{W}_t\,(\boldsymbol{h}_t(\boldsymbol{\gamma}) - \boldsymbol{h}_t(\boldsymbol{\gamma}_0))\right\}\right) > 0.$$

Since $\{\boldsymbol{h}_t(\boldsymbol{\gamma}_0), \mathcal{F}_t\}$ is a martingale difference sequence, and $\boldsymbol{W}_t$ is measurable–$\mathcal{F}_{t-1}$, we have $E\{\boldsymbol{h}_t'(\boldsymbol{\gamma})\boldsymbol{W}_t\boldsymbol{h}_t(\boldsymbol{\gamma})\} = E\left\{(\boldsymbol{h}_t(\boldsymbol{\gamma}) - \boldsymbol{h}_t(\boldsymbol{\gamma}_0))'\,\boldsymbol{W}_t\,(\boldsymbol{h}_t(\boldsymbol{\gamma}) - \boldsymbol{h}_t(\boldsymbol{\gamma}_0))\right\} + E\{\boldsymbol{h}_t'(\boldsymbol{\gamma}_0)\boldsymbol{W}_t\boldsymbol{h}_t(\boldsymbol{\gamma}_0)\}$. Since $\boldsymbol{W}_t$ is non-negative definite a.s.-P, assumption 3 ensures that the uniqueness of the minimum of $\overline{Q}_T(\boldsymbol{\gamma})$ does not vanish as T becomes arbitrary large. If the process $\left\{(\boldsymbol{x}_t', y_t)'\right\}$ is stationary, $f_t = f : \mathbb{R}^{\upsilon} \times \Theta \to \mathbb{R}^1$, and $\boldsymbol{W}_t = \boldsymbol{W}\left(\boldsymbol{v}_{t-p}', \ldots, \boldsymbol{v}_t'\right)$ is positive definite a.s.-P, then assumption 3 is equivalent to say that: $f(\cdot, \boldsymbol{\theta}) = f(\cdot, \boldsymbol{\theta}_0)$ a.s.-P only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

By applying theorem 3.4 of White (1996) we obtain the following result.

Result 1 *Under assumptions 1–3, $\hat{\boldsymbol{\gamma}}_T \xrightarrow{a.s.} \boldsymbol{\gamma}_0$, as $T \to \infty$.*

The following assumptions are sufficient to study the asymptotic distribution of our estimator.

[1] A sequence of random variables $\{D_t\}$ is uniformly L_r-bounded if $\sup_t E\left|D_t\right|^r < \infty$.

Assumption 4 *The point* $\boldsymbol{\gamma}_0$ *is interior to* Γ.

Assumption 5 *The random functions* $f_t(\cdot, \boldsymbol{\theta})$ *are continuously differentiable of order 2 on* Γ *uniformly in* t *a.s.-P.*

Assumption 6

$$\sup_{t\in\mathbb{N}} E\left\{\|\boldsymbol{W}_t\| \sup_{\Theta}\left(\left\|\nabla^2_{\boldsymbol{\theta}} f_t(\cdot,\boldsymbol{\theta})\right\|^2 + \sum_{i=0}^{p}\|\nabla_{\boldsymbol{\theta}} f_{t-i}(\cdot,\boldsymbol{\theta})\|^4 + \sum_{i=1}^{p}\epsilon^2_{t-i}\left\|\nabla^2_{\boldsymbol{\theta}} f_{t-i}(\cdot,\boldsymbol{\theta})\right\|^2 \right.\right.$$
$$\left.\left. + \sum_{i=0}^{p} f^2_{t-i}(\cdot,\boldsymbol{\theta})\left\|\nabla^2_{\boldsymbol{\theta}} f_{t-i}(\cdot,\boldsymbol{\theta})\right\|^2\right)\right\}^r < \infty, \text{ for some } r > \frac{a}{a-1}.$$

Assumption 7 *The sequence* $\left\{\bar{\boldsymbol{A}}_T(\boldsymbol{\gamma}_0) = 2T^{-1}\sum_{t=1}^{T} E\left\{\nabla_{\boldsymbol{\gamma}}\boldsymbol{h}'_t(\boldsymbol{\gamma}_0)\boldsymbol{W}_t\nabla_{\boldsymbol{\gamma}'}\boldsymbol{h}_t(\boldsymbol{\gamma}_0)\right\}\right\}$ *is* $\boldsymbol{O}(1)$, *and* $\liminf_{T\to\infty}\left|\bar{\boldsymbol{A}}_T(\boldsymbol{\gamma}_0)\right| > 0$.

Assumption 8 *For some* $r > \dfrac{a}{a-1}$,

$$\sup_{t\in\mathbb{N}} E\left\{\|\boldsymbol{W}_t\|^2\left(1 + f^2_t(\cdot,\boldsymbol{\theta}_0)\|\nabla_{\boldsymbol{\theta}} f_t(\cdot,\boldsymbol{\theta}_0)\|^2 + \sum_{i=1}^{p}\epsilon^2_{t-i}\|\nabla_{\boldsymbol{\theta}} f_{t-i}(\cdot,\boldsymbol{\theta}_0)\|^2\right.\right.$$
$$\left.\left. + \|\nabla_{\boldsymbol{\theta}} f_t(\cdot,\boldsymbol{\theta}_0)\|^2 + \sum_{i=1}^{p}\epsilon^4_{t-i}\right)\left(1 + \sum_{i=0}^{p}\epsilon^4_{t-i} + \epsilon^2_t f^2_t(\cdot,\boldsymbol{\theta}_0)\right)\right\}^r < \infty.$$

Both of assumptions 2, 6 and 8 are sufficient for general cases and can be replaced by much simpler conditions for specific choice of $\boldsymbol{W}_t$ as shown at the end of section 1.3.

Assumption 9 *The sequence*

$$\left\{ \boldsymbol{V}_T = 4T^{-1} \sum_{t=1}^{T} E\left\{ \nabla_{\gamma} \boldsymbol{h}_t'(\boldsymbol{\gamma}_0) \boldsymbol{W}_t \boldsymbol{h}_t(\boldsymbol{\gamma}_0) \boldsymbol{h}_t'(\boldsymbol{\gamma}_0) \boldsymbol{W}_t \nabla_{\gamma'} \boldsymbol{h}_t(\boldsymbol{\gamma}_0) \right\} \right\}$$

is $\boldsymbol{O}(1)$, *and* $\liminf_{T\to\infty} |\boldsymbol{V}_T| > 0$.

The following theorem gives the limiting distribution of a scaled version of our estimator. The proof is given in appendix A.

Theorem 1 *Given assumptions 1–9 we have*

$$\boldsymbol{V}_T^{-1/2} \bar{\boldsymbol{A}}_T(\boldsymbol{\gamma}_0) \sqrt{T} \left(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_0\right) \xrightarrow{d} N\left(\boldsymbol{0}, \boldsymbol{I}_{q+p+1}\right) \quad \text{as } T \to \infty.$$

This result is true for any member of the class of estimators defined by equation (1.3) if $\boldsymbol{W}_t$ satisfy assumptions 2, 3, 6–9. In the following section we derive the most efficient estimator among this class of estimators.

1.3 Asymptotically Optimal SLS

The asymptotic covariance (acov) of $\sqrt{T}\,(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_0)$ is given by $\bar{\boldsymbol{A}}_T^{-1}(\boldsymbol{\gamma}_0)\boldsymbol{V}_T\bar{\boldsymbol{A}}_T^{-1}(\boldsymbol{\gamma}_0)$ which depends on $\boldsymbol{W}_t$, $t = 1, 2, \ldots, T$. A reasonable definition of the asymptotically optimal estimator in the class defined by equation 1.3 is the one which minimizes the asymptotic variance of $\sqrt{T}\,\boldsymbol{a}'\left(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_0\right)$, $\boldsymbol{a} \in \mathbb{R}^{q+p+1}$. The following proposition describes the optimal choice of $\boldsymbol{W}_t$ which achieves this criterion. The proof is provided in the appendix A.

Proposition 2 *Suppose that* $\boldsymbol{U}_t = E\left\{\boldsymbol{h}_t(\boldsymbol{\gamma}_0)\boldsymbol{h}_t'(\boldsymbol{\gamma}_0) \mid \boldsymbol{v}_{t-p}', \ldots, \boldsymbol{v}_t'\right\}$ *is nonsingular a.s.-P, and assumptions 2, 3, 6–9 hold with* $\boldsymbol{W}_t = \boldsymbol{U}_t^{-1}$. *Then the asymptotically optimal SLS estimator;* $\hat{\boldsymbol{\gamma}}_T^o$, *in the class of estimators defined by equation (1.3) is obtained by letting* $\boldsymbol{W}_t = \boldsymbol{U}_t^{-1}$, $t = 1, 2, \ldots, T$, *and we have*

$$acov^{-1}\sqrt{T}\ (\hat{\boldsymbol{\gamma}}_T^o - \boldsymbol{\gamma}_0) = T^{-1}\sum_{t=1}^{T} E\left\{\nabla_{\boldsymbol{\gamma}}\boldsymbol{h}_t'(\boldsymbol{\gamma}_0)\boldsymbol{U}_t^{-1}\nabla_{\boldsymbol{\gamma}'}\boldsymbol{h}_t(\boldsymbol{\gamma}_0)\right\}, \tag{1.4}$$

or equivalently,

$$acov^{-1}\sqrt{T}\ (\hat{\boldsymbol{\gamma}}_T^o - \boldsymbol{\gamma}_0) = T^{-1}\sum_{t=1}^{T} E\left\{\boldsymbol{B}_t'\boldsymbol{\Omega}_t^{-1}\boldsymbol{B}_t\right\}, \tag{1.5}$$

where

$$\boldsymbol{B}_t' = \begin{pmatrix} \nabla_{\boldsymbol{\theta}} f_t\left(\boldsymbol{v}_t, \boldsymbol{\theta}_0\right) & \nabla_{\boldsymbol{\theta}}\sigma_t^2(\boldsymbol{\gamma}_0) \\ \boldsymbol{0} & \nabla_{\boldsymbol{\phi}}\sigma_t^2(\boldsymbol{\gamma}_0) \end{pmatrix}, \quad \textit{and} \tag{1.6}$$

$$\boldsymbol{\Omega}_t = \sigma_t^2(\boldsymbol{\gamma}_0)\begin{pmatrix} 1 & \sigma_t(\boldsymbol{\gamma}_0)E\left(\varepsilon_t^3 \mid \boldsymbol{v}_{t-p}', \ldots, \boldsymbol{v}_t'\right) \\ \cdot & \sigma_t^2(\boldsymbol{\gamma}_0)\left[E\left(\varepsilon_t^4 \mid \boldsymbol{v}_{t-p}', \ldots, \boldsymbol{v}_t'\right) - 1\right] \end{pmatrix}. \tag{1.7}$$

It is clear that $\boldsymbol{U}_t$ depends on $\boldsymbol{\gamma}_0$, $E\left(\varepsilon_t^3 \mid \boldsymbol{v}_{t-p}', \ldots, \boldsymbol{v}_t'\right)$, and $E\left(\varepsilon_t^4 \mid \boldsymbol{v}_{t-p}', \ldots, \boldsymbol{v}_t'\right)$. Therefore, $\hat{\boldsymbol{\gamma}}_T^o$ is infeasible. Two-step estimator can be calculated as follow. First, a preliminary consistent estimator of $\boldsymbol{\gamma}_0$ is calculated, such as the QMLE or simply $\hat{\boldsymbol{\gamma}}_T$ based on the identity weight matrix. Second, the residuals $\hat{\varepsilon}_t$ are calculated, then suitable AR models are fitted to $\hat{\varepsilon}_t^3$ and $\hat{\varepsilon}_t^4$ respectively. Finally, replace $\boldsymbol{\gamma}_0$, $E\left(\varepsilon_t^3 \mid \boldsymbol{v}_{t-p}', \ldots, \boldsymbol{v}_t'\right)$, and $E\left(\varepsilon_t^4 \mid \boldsymbol{v}_{t-p}', \ldots, \boldsymbol{v}_t'\right)$ in $\boldsymbol{U}_t$ by the corresponding fitted values and use $\boldsymbol{W}_t = \widehat{\boldsymbol{U}}_t^{-1}$ in equation (1.3). Fitting the AR models may be useful if the driving noise sequence $\{\varepsilon_t\}$ is not i.i.d., otherwise it is enough to use the sample means of $\hat{\varepsilon}_t^3$ and $\hat{\varepsilon}_t^4$ respectively. Under fairly general conditions the resulting two-step estimator (henceforth FOSLS) is consistent even if the preliminary sample statistics (used in the first step) lead to inconsistent estimator for $\boldsymbol{U}_t$. Moreover, if $\widehat{\boldsymbol{U}}_t$ is consistent for $\boldsymbol{U}_t$, the FOSLS estimator does have the same asymptotic variance given by (1.4). For more details about the asymptotics of two-step estimators the reader is referred to (White, 1996, section 6.3). Alternatively, if the conditioning set $\left\{\boldsymbol{v}'_{t-p}, \ldots, \boldsymbol{v}'_t\right\}$ is reasonably small, we can use nonparametric regression models of the conditional skewness and kurtosis to obtain $\widehat{\boldsymbol{U}}_t$.

Before we conclude this section, it is worthwhile to reconsider assumptions 2, 6 and 8 in case of using the optimal weight matrix. After long steps of mathematical simplifications which involve using Minkowski inequality and c_r inequality, it appears that we can replace assumptions 2, 6 and 8 by the following three assumptions respectively if $\boldsymbol{W}_t = \boldsymbol{U}_t^{-1}$

Assumption 10 *For $k = 0, 1, \ldots,$*

$$\sup_{t\in\mathbb{N}} E\left\{\frac{\epsilon_{t-k}^4 + \sup_\Theta f_{t-k}^4\left(\cdot, \boldsymbol{\theta}\right)}{\sigma_t^4}\right\}^r < \infty, \textit{ for some } r > \frac{a}{a-1}.$$

Assumption 11 *For $s = 1, 2,$ and $k = 0, 1, \ldots,$*

$$\sup_{t\in\mathbb{N}} E\left\{\frac{\sup_\Theta \left\|\nabla_{\boldsymbol{\theta}}^s f_{t-k}\left(\cdot, \boldsymbol{\theta}\right)\right\|^4}{\sigma_t^4}\right\}^r < \infty, \textit{ for some } r > \frac{a}{a-1}.$$

Assumption 12 *For* $k = 0, 1, \ldots,$

$$\sup_{t \in \mathbb{N}} E \left\{ \frac{f_t^8(\cdot, \boldsymbol{\theta}_0) + \epsilon_{t-k}^8 + \left\| \nabla_{\boldsymbol{\theta}} f_{t-k}(\cdot, \boldsymbol{\theta}_0) \right\|^8}{\sigma_t^8} \right\}^r < \infty, \textit{ for some } r > \frac{a}{a-1}.$$

It is important here to emphasize that these alternative assumptions are sufficient but not necessary to prove Result 1 and Theorem 1. As we shall see from the Monte Carlo results (Table 1.2) the small/moderate/large sample properties of the optimal SLSE are not affected even if the innovation's moments higher than four don't exist. Actually, it is not unexpected to see the asymptotic theory of the optimal SLS working well under the same regularity conditions of the QML theory. This is most likely due to the fact that the optimal SLS is a kind of refinement to the QML using the information inherent in the skewness and kurtosis of the process innovation.

In the following section we investigate the gain of efficiency of the optimal SLS estimator compared to the Gaussian quasi-ML and the Estimating Function (EF) estimators.

1.4 Relative Efficiency of the Optimal SLSE

The Gaussian QMLE is considered to be the most popular method of estimation in the dynamic econometric models that jointly parametrize conditional means and conditional variances. It is generally obtained by maximizing a normal log-likelihood. The asymptotic properties of the Gaussian QMLE were firstly studied by Weiss (1986) for univariate ARMA model with ARCH disturbances. Latter, his results were extended to multivariate GARCH models in (Bollerslev and Wooldridge, 1992).

Specifically for model (1.1) - (1.2), the global Gaussian QMLE is defined by functions $\hat{\boldsymbol{\gamma}}_T^Q : \Omega \to \Gamma$ measurable–$\mathcal{F}$, $T = 1, 2, \ldots$ such that

$$\hat{\boldsymbol{\gamma}}_T^Q = \underset{\boldsymbol{\gamma} \in \Gamma}{\operatorname{argmin}}\, T^{-1} \sum_{t=1}^{T} \log \sigma_t^2(\boldsymbol{\gamma}) + \frac{\epsilon_t^2(\boldsymbol{\theta})}{\sigma_t^2(\boldsymbol{\gamma})} \qquad \text{a.s.-}P. \tag{1.8}$$

Under adapted version of assumptions 2–9, we can follow the same steps found in section 1.2 and the proof of theorem 1 to show that $\hat{\boldsymbol{\gamma}}_T^Q$ is $\sqrt{T}$-consistent estimator with $\text{acov}\sqrt{T}\left(\hat{\boldsymbol{\gamma}}_T^Q - \boldsymbol{\gamma}_0\right)$ given by

$$T\left(\sum_{t=1}^{T} E\left\{\boldsymbol{B}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{B}_t\right\}\right)^{-1}\left(\sum_{t=1}^{T} E\left\{\boldsymbol{B}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\Omega}_t\boldsymbol{\Sigma}_t^{-1}\boldsymbol{B}_t\right\}\right)\left(\sum_{t=1}^{T} E\left\{\boldsymbol{B}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{B}_t\right\}\right)^{-1}, \tag{1.9}$$

where $\boldsymbol{B}_t$, $\boldsymbol{\Omega}_t$ are defined by (1.6) and (1.7) respectively, and

$$\boldsymbol{\Sigma}_t = \begin{pmatrix} \sigma_t^2(\boldsymbol{\gamma}_0) & 0 \\ 0 & 2\sigma_t^4(\boldsymbol{\gamma}_0) \end{pmatrix}.$$

Hence, by using argument similar to that used in the proof of proposition 2, we can show that

$$\text{acov}\sqrt{T}\,\boldsymbol{a}'\left(\hat{\boldsymbol{\gamma}}_T^o - \boldsymbol{\gamma}_0\right) \le \text{acov}\sqrt{T}\,\boldsymbol{a}'\left(\hat{\boldsymbol{\gamma}}_T^Q - \boldsymbol{\gamma}_0\right), \quad \boldsymbol{a} \in \mathbb{R}^{q+p+1},$$

and for a given $\boldsymbol{a} \in \mathbb{R}^{q+p+1}$, the equality holds if and only if for $t = 1, 2, \ldots, T$,

$$\boldsymbol{\Omega}_t\boldsymbol{\Sigma}_t^{-1}\boldsymbol{B}_t\boldsymbol{a} = \boldsymbol{B}_t\left(\sum_{t=1}^{T} E\left\{\boldsymbol{B}_t'\boldsymbol{\Omega}_t^{-1}\boldsymbol{B}_t\right\}\right)^{-1}\left(\sum_{t=1}^{T} E\left\{\boldsymbol{B}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{B}_t\right\}\right)\boldsymbol{a} \quad \text{a.s.-}P. \tag{1.10}$$

This n.s. condition is quite general and can be simplified under specific settings. For example, if the process $\left\{(\boldsymbol{x}_t', y_t, \sigma_t, \epsilon_t)'\right\}$ is stationary with $E\left(\varepsilon_t^3 \mid \boldsymbol{v}_{t-p}', \ldots, \boldsymbol{v}_t'\right) = 0$, and $E\left(\varepsilon_t^4 \mid \boldsymbol{v}_{t-p}', \ldots, \boldsymbol{v}_t'\right) = \mu_4$, then it can be shown that equation (1.10) is equivalent to

$$\boldsymbol{a}_1'\left(\boldsymbol{I}_q - \boldsymbol{C}\right)\nabla_{\boldsymbol{\theta}} f_t\left(\boldsymbol{v}_t, \boldsymbol{\theta}_0\right) = 0, \quad \boldsymbol{a}_1'\left(\frac{\mu_4 - 1}{2}\boldsymbol{I}_q - \boldsymbol{C}\right)\nabla_{\boldsymbol{\theta}}\sigma_t^2(\boldsymbol{\gamma}_0) = 0, \tag{1.11}$$

where

$\boldsymbol{C} = \left(\boldsymbol{C}_1 + \frac{1}{2}\boldsymbol{C}_2\right)\left(\boldsymbol{C}_1 + \frac{1}{\mu_4 - 1}\boldsymbol{C}_2\right)^{-1}$, $\boldsymbol{C}_1 = E\left\{\sigma_t^{-2}(\boldsymbol{\gamma}_0)\nabla_{\boldsymbol{\theta}} f_t\left(\boldsymbol{v}_t, \boldsymbol{\theta}_0\right)\nabla_{\boldsymbol{\theta}'} f_t\left(\boldsymbol{v}_t, \boldsymbol{\theta}_0\right)\right\}$, $\boldsymbol{C}_2 = E\left\{\sigma_t^{-4}(\boldsymbol{\gamma}_0)\nabla_{\boldsymbol{\theta}}\sigma_t^2(\boldsymbol{\gamma}_0)\nabla_{\boldsymbol{\theta}'}\sigma_t^2(\boldsymbol{\gamma}_0)\right\}$, and $\boldsymbol{a}_1$ is a sub-vector which contains the first q elements of $\boldsymbol{a}$.

The EF is a general alternative approach which makes use of the available information inherent in the conditional skewness and kurtosis of the process to improve the efficiency of the QMLE. This makes it comparable to our approach. In light of the Corollary given by Durairajan (1992), it can be shown that the EF estimator; $\hat{\boldsymbol{\gamma}}_T^{EF}$, under model (1.1) - (1.2) is obtained by solving the equation

$$\boldsymbol{g}^*(\boldsymbol{\gamma}) = \sum_{t=1}^{T}\boldsymbol{B}_t'(\boldsymbol{\gamma})\boldsymbol{\Omega}_t^{-1}(\boldsymbol{\gamma})\begin{pmatrix}\epsilon_t(\boldsymbol{\gamma}) \\ \epsilon_t^2(\boldsymbol{\gamma}) - \sigma_t^2(\boldsymbol{\gamma})\end{pmatrix} = \mathbf{0}, \tag{1.12}$$

where $\boldsymbol{B}_t'(\boldsymbol{\gamma})$, and $\boldsymbol{\Omega}_t(\boldsymbol{\gamma})$ are given by equations (1.6) and (1.7) after replacing $\boldsymbol{\theta}_0$, $\boldsymbol{\gamma}_0$, $E\left(\varepsilon_t^3 \mid \boldsymbol{v}_{t-p}', \ldots, \boldsymbol{v}_t'\right)$, and $E\left(\varepsilon_t^4 \mid \boldsymbol{v}_{t-p}', \ldots, \boldsymbol{v}_t'\right)$ with $\boldsymbol{\theta}$, $\boldsymbol{\gamma}$, $E_{\boldsymbol{\gamma}}\left(\varepsilon_t^3(\boldsymbol{\gamma}) \mid \boldsymbol{v}_{t-p}', \ldots, \boldsymbol{v}_t'\right)$, and $E_{\boldsymbol{\gamma}}\left(\varepsilon_t^4(\boldsymbol{\gamma}) \mid \boldsymbol{v}_{t-p}', \ldots, \boldsymbol{v}_t'\right)$ respectively. The function $\boldsymbol{g}^*(\boldsymbol{\gamma})$ is optimal with respect to Godambe's information criterion. That is, consider the class of estimating functions $\mathcal{G} = \{\boldsymbol{g}(\boldsymbol{\gamma})\}$, such that

$$\boldsymbol{g}(\boldsymbol{\gamma}) = \sum_{t=1}^{T} \boldsymbol{K}'_t(\boldsymbol{\gamma}) \left(\begin{array}{c} \epsilon_t(\boldsymbol{\gamma}) \\ \epsilon_t^2(\boldsymbol{\gamma}) - \sigma_t^2(\boldsymbol{\gamma}) \end{array} \right),$$

where $\boldsymbol{K}'_t(\boldsymbol{\gamma})$ is a $(q + p + 1) \times 2$ matrix which is measurable with respect to $\mathcal{F}\left\{\boldsymbol{v}'_{t-p}, \ldots, \boldsymbol{v}'_t\right\}$ and full column rank a.s.-P, and both of $E_{\boldsymbol{\gamma}}\left\{\boldsymbol{g}(\boldsymbol{\gamma})\boldsymbol{g}'(\boldsymbol{\gamma})\right\}$ and $E_{\boldsymbol{\gamma}}\left\{\nabla_{\boldsymbol{\gamma}'}\boldsymbol{g}(\boldsymbol{\gamma})\right\}$ are nonsingular for each $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$. Then, it can be directly shown that for every $\boldsymbol{a} \in \mathbb{R}^{q+p+1}$, $\boldsymbol{g}^*(\boldsymbol{\gamma})$ is an optimal choice which minimizes

$$\boldsymbol{a}' \left(E_{\boldsymbol{\gamma}}^{-1}\left\{\nabla_{\boldsymbol{\gamma}'}\boldsymbol{g}(\boldsymbol{\gamma})\right\} E_{\boldsymbol{\gamma}}\left\{\boldsymbol{g}(\boldsymbol{\gamma})\boldsymbol{g}'(\boldsymbol{\gamma})\right\} E_{\boldsymbol{\gamma}}^{-1}\left\{\nabla_{\boldsymbol{\gamma}}\boldsymbol{g}'(\boldsymbol{\gamma})\right\}\right) \boldsymbol{a}.$$

Under some regularity conditions similar to assumptions 2–9, we can show that $\hat{\boldsymbol{\gamma}}_T^{EF}$ is a $\sqrt{T}$-consistent estimator with acov$\sqrt{T}\left(\hat{\boldsymbol{\gamma}}_T^{EF} - \boldsymbol{\gamma}_0\right)$ given by equation (1.5). Although both of $\hat{\boldsymbol{\gamma}}_T^{o}$ and $\hat{\boldsymbol{\gamma}}_T^{EF}$ share the same asymptotic variance, but they are distinct due to the following reasons. First, the FOSLS is an extremum estimator compared to the EF estimator which represents a solution of the optimal estimating equation (1.12). Second, if $E_{\boldsymbol{\gamma}}\left(\varepsilon_t^3(\boldsymbol{\gamma}) \mid \boldsymbol{v}'_{t-p}, \ldots, \boldsymbol{v}'_t\right)$, and $E_{\boldsymbol{\gamma}}\left(\varepsilon_t^4(\boldsymbol{\gamma}) \mid \boldsymbol{v}'_{t-p}, \ldots, \boldsymbol{v}'_t\right)$ are completely known functions of $\boldsymbol{\gamma}$, then $\hat{\boldsymbol{\gamma}}_T^{EF}$ is clearly a one-step estimator compared to $\hat{\boldsymbol{\gamma}}_T^{o}$ which remains infeasible due the dependence of $\boldsymbol{W}_t$ on $\boldsymbol{\gamma}_0$. Third, the two estimators may have different finite-sample behaviour. This can be seen clearly by comparing equation (1.12) with the first order condition for $\hat{\boldsymbol{\gamma}}_T^{o}$ which can be written as

$$\sum_{t=1}^{T} \boldsymbol{B}'_t(\boldsymbol{\gamma})\boldsymbol{H}_t(\boldsymbol{\theta})\boldsymbol{\Omega}_t^{-1}\boldsymbol{H}'_t(\boldsymbol{\theta}) \left(\begin{array}{c} \epsilon_t(\boldsymbol{\gamma}) \\ \epsilon_t^2(\boldsymbol{\gamma}) - \sigma_t^2(\boldsymbol{\gamma}) \end{array} \right) = \boldsymbol{0},$$

where

$$\boldsymbol{H}_t(\boldsymbol{\theta}) = \begin{pmatrix} 1 & 2f_t\left(\boldsymbol{v}_t, \boldsymbol{\theta}\right) - 2f_t\left(\boldsymbol{v}_t, \boldsymbol{\theta}_0\right) \\ 0 & 1 \end{pmatrix}.$$

We conclude this section by examining how much loss of efficiency in the QML estimator is recovered by using the OSLS estimator instead. Unfortunately, a general answer to this question is not possible in light of equations (1.5) and (1.9). However, it is straightforward to investigate this issue for special cases of model (1.1) - (1.2). For convenience, we consider AR(1) model with ARCH(1) disturbances. The model is given by $y_t = \theta_0 y_{t-1} + \epsilon_t$, $\sigma_t^2 = 1 - \phi_0 + \phi_0 \epsilon_{t-1}^2$, and the innovations sequence $\{\varepsilon_t = \epsilon_t / \sigma_t\}$ is i.i.d. drawn from a standardized[2] distribution. This model has two dynamic components represented by the parameters θ_0 and ϕ_0 in the conditional mean and variance functions respectively. Since the OSLS estimator makes use of the conditional skewness and kurtosis of $\{\varepsilon_t\}$, we consider both symmetric and skewed distributions in our calculations. The asymptotic variance of the true ML estimator is calculated as benchmark.

Table 1.1 (page 23) gives a snapshot of the obtained results. It shows that the out-performance of the OSLS estimator over the QML is more emphasized in case of highly skewed distributions such as the Gamma distribution. On the other hand, the out-performance of the OSLS estimator diminishes in case of symmetric distributions such as the Student-t distribution. This is obviously seen by comparing the asymptotic variances of the QML and the OSLS estimator of ϕ_0 for any pattern under the Student-t distribution. They turn out to be equal, which is consistent with the theoretical result given by (1.11).

[2] A standardized distribution refers to a distribution with zero mean and unit variance.

Table 1.1: Asymptotic variances of the OSLS, QML, and ML estimators

Distribution	Estimation approach	$v(\hat{\theta}_0)$ Pattern (a)	$v(\hat{\phi}_0)$	$v(\hat{\theta}_0)$ Pattern (b)	$v(\hat{\phi}_0)$	$v(\hat{\theta}_0)$ Pattern (c)	$v(\hat{\phi}_0)$	$v(\hat{\theta}_0)$ Pattern (d)	$v(\hat{\phi}_0)$
		$\theta_0 = 0.2$ $\phi_0 = 0.2$		$\theta_0 = 0.2$ $\phi_0 = 0.6$		$\theta_0 = 0.8$ $\phi_0 = 0.2$		$\theta_0 = 0.8$ $\phi_0 = 0.6$	
Student-t (5)	OSLS	1.34	6.32	1.51	2.86	0.37	6.26	0.29	2.86
	QML	1.56	6.32	2.34	2.86	0.41	6.26	0.41	2.86
	ML	1.05	2.44	1.04	1.11	0.29	2.41	0.21	1.11
Student-t (7)	OSLS	1.26	3.30	1.26	1.51	0.37	3.32	0.29	1.51
	QML	1.30	3.30	1.41	1.51	0.38	3.32	0.32	1.51
	ML	1.10	2.31	1.05	1.05	0.32	2.34	0.25	1.05
Student-t (13)	OSLS	1.20	2.31	1.08	1.06	0.37	2.33	0.26	1.06
	QML	1.20	2.31	1.10	1.06	0.37	2.33	0.26	1.06
	ML	1.15	2.12	1.03	0.97	0.36	2.17	0.24	0.97
Gamma (2)	OSLS	0.83	2.89	0.81	1.23	0.24	2.77	0.19	1.23
	QML	1.35	4.48	1.63	2.02	0.42	4.48	0.38	2.03
	ML	0.15	0.25	0.08	0.14	0.04	0.06	0.04	0.13
Gamma (8)	OSLS	0.97	2.08	0.88	0.94	0.31	2.06	0.22	0.94
	QML	1.18	2.52	1.09	1.15	0.38	2.51	0.28	1.14
	ML	0.87	1.44	0.69	0.64	0.27	1.35	0.19	0.62
Gamma (12)	OSLS	1.02	2.00	0.90	0.90	0.32	1.99	0.23	0.90
	QML	1.17	2.30	1.05	1.04	0.37	2.29	0.27	1.04
	ML	0.97	1.58	0.79	0.69	0.30	1.50	0.21	0.72
Gamma (20)	OSLS	1.06	1.93	0.91	0.88	0.34	1.93	0.24	0.88
	QML	1.15	2.11	1.00	0.96	0.37	2.11	0.26	0.96
	ML	1.03	1.67	0.86	0.73	0.33	1.62	0.23	0.75

Model: $y_t = \theta_0 y_{t-1} + \epsilon_t$, $\sigma_t^2 = 1 - \phi_0 + \phi_0 \epsilon_{t-1}^2$, and ϵ_t/σ_t are drawn independently from the standardized version of the listed distributions. The numbers in the brackets represent the shape parameters of the Gamma distributon and the degrees of freedom for the T distribution.

In order to study how much loss of efficiency in the QML estimator is recovered by the OSLS estimator, we suggest the following measure of reduction in the QML efficiency-loss (inefficiency);

$$RIFL\left(\boldsymbol{a}'\boldsymbol{\gamma}_0\right) = 100 * \frac{\text{acov}\sqrt{T}\,\boldsymbol{a}'\left(\hat{\boldsymbol{\gamma}}_T^Q - \boldsymbol{\gamma}_0\right) - \text{acov}\sqrt{T}\,\boldsymbol{a}'\left(\hat{\boldsymbol{\gamma}}_T^o - \boldsymbol{\gamma}_0\right)}{\text{acov}\sqrt{T}\,\boldsymbol{a}'\left(\hat{\boldsymbol{\gamma}}_T^Q - \boldsymbol{\gamma}_0\right) - \text{acov}\sqrt{T}\,\boldsymbol{a}'\left(\hat{\boldsymbol{\gamma}}_T^M - \boldsymbol{\gamma}_0\right)},$$

where $\hat{\boldsymbol{\gamma}}_T^M$ is the ML estimator of $\boldsymbol{\gamma}_0$. This measure can be also used to examine which estimator approaches to the asymptotic variance lower bound faster.

Figure 1.1 (page 25) is produced as a sample of output. According to Figure 1.1-(a), if the driving noise has a heavy tail Student-t distribution, 45–60% of the inefficiency of $\hat{\theta}_T^Q$ is recovered by $\hat{\theta}_T^o$. The $RIFL(\theta_0)$ declines sharply as the degrees of freedom get higher. This indicates that as the symmetric distribution gets close to the Gaussian distribution, the performance of the QML is improved very fast and it becomes very close to the OSLS estimator, and both of them approaches the variance lower bound. The situation is opposite in Figure 1.1-(b,c) which show that $RIFL(\theta_0)$ and $RIFL(\phi_0)$ are increasing functions of the shape parameter of the Gamma distribution. This indicates that as the skewed distribution gets closer to the Gaussian distribution the performance of the OSLS estimator is improved significantly faster than the performance of the QML. In other words the QML is persistently affected by the conditional skewness and it is preferable to use the optimal estimator. Clearly all of these results apply if the OSLS is replaced by the EF estimator. In the next section we investigate the finite-sample performance of the FOSLS in comparison to the QML and EF estimators.

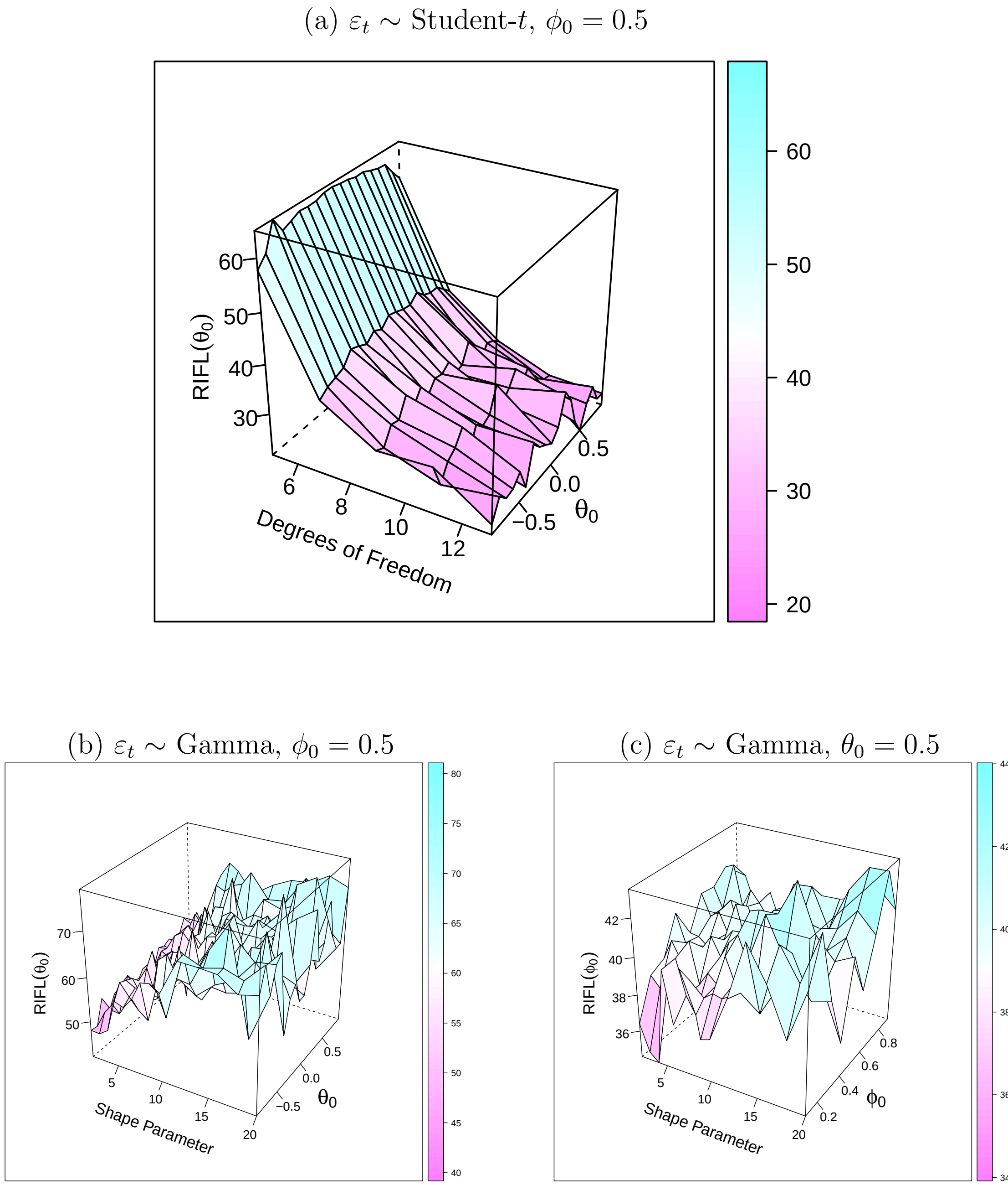


Figure 1.1: RIFL is Reduction (%) in the QML inefficiency

1.5 Monte Carlo Simulations

In this section we run several Monte Carlo simulations to present some of the finite-sample properties of the FOSLS estimator relative to both the QML and EF estimators. For the sake of simplicity, we generate the data from the AR(1)-ARCH(1) model which was introduced in section 1.4. The length of the time series; T; varies from 30 to 1000. We also consider $T = 10,000$ to establish a link with the asymptotic results in section 1.4. In each simulation the autoregressive parameters (θ_0, ϕ_0) are varied to represent different levels of persistence in the mean and variance equations. The innovations sequence is drawn from standardized distributions having different levels of skewness and kurtosis. The comparisons are based on two common finite-sample criteria, namely the bias and the root mean squared error (RMSE).

Tables (1.2, 1.3) (pages 27,28) report the sample means $(\bar{\hat{\theta}}_0, \bar{\hat{\phi}}_0)$, and root mean squared errors $(RMSE_{\hat{\theta}_0}, RMSE_{\hat{\phi}_0})$ for each estimator based on 3000 independent replications over four different pairs of (θ_0, ϕ_0) and three sample sizes.

Table 1.2 presents summary results for Student-t innovations with 5 df. It shows slight gain of efficiency in estimating θ_0 if the FOSLS or EF estimators are used instead of the QML specially for high value of ϕ_0. Otherwise, the QML performs fairly well in small samples. It is interesting to see the three estimator having the same degree of bias almost everywhere.

Table 1.3 presents summary results for Gamma(2,1) innovations. The over-performance of the FOSLS and EF relative to the QML is clear for all sample sizes and panels. It is also clear that the three estimator have the same degree of bias

Table 1.2: Finite-sample properties of QML, EF, and FOSLS under Student-$t(5)$

T	Estimator	$\bar{\hat{\theta}}_0$	$RMSE_{\hat{\theta}_0}$	$\bar{\hat{\phi}}_0$	$RMSE_{\hat{\phi}_0}$	$\bar{\hat{\theta}}_0$	$RMSE_{\hat{\theta}_0}$	$\bar{\hat{\phi}}_0$	$RMSE_{\hat{\phi}_0}$
		Panel (a): $\theta_0 = \phi_0 = 0.2$				Panel (b): $\theta_0 = 0.2$, $\phi_0 = 0.6$			
60	QML	0.19	0.162	0.26	0.184	0.19	0.162	0.53	0.190
	EF	0.19	0.153	0.26	0.177	0.19	0.159	0.53	0.193
	FOSLS	0.19	0.157	0.27	0.189	0.19	0.159	0.53	0.195
100	QML	0.19	0.130	0.22	0.151	0.20	0.129	0.53	0.159
	EF	0.19	0.122	0.22	0.148	0.20	0.124	0.53	0.167
	FOSLS	0.19	0.123	0.22	0.150	0.20	0.125	0.53	0.166
1000	QML	0.20	0.040	0.18	0.068	0.20	0.044	0.58	0.050
	EF	0.20	0.037	0.18	0.068	0.20	0.039	0.57	0.050
	FOSLS	0.20	0.037	0.18	0.068	0.20	0.039	0.57	0.050
		Panel (c): $\theta_0 = 0.8$, $\phi_0 = 0.2$				Panel (d): $\theta_0 = 0.8$, $\phi_0 = 0.6$			
60	QML	0.77	0.099	0.26	0.188	0.78	0.088	0.53	0.190
	EF	0.77	0.093	0.26	0.182	0.78	0.099	0.52	0.206
	FOSLS	0.77	0.094	0.27	0.192	0.78	0.085	0.54	0.194
100	QML	0.78	0.074	0.22	0.154	0.78	0.070	0.54	0.156
	EF	0.78	0.071	0.23	0.151	0.78	0.078	0.53	0.165
	OLS	0.78	0.070	0.22	0.154	0.78	0.065	0.54	0.160
1000	QML	0.80	0.021	0.18	0.068	0.80	0.021	0.57	0.051
	EF	0.80	0.020	0.18	0.068	0.80	0.018	0.57	0.052
	FOSLS	0.80	0.020	0.18	0.068	0.80	0.018	0.57	0.051

Model: $y_t = \theta_0 y_{t-1} + \epsilon_t$, $\sigma_t^2 = 1 - \phi_0 + \phi_0 \epsilon_{t-1}^2$, and ϵ_t/σ_t are drawn independently from the standardized Student-t distribution with five degrees of freedom. The sample mean and root mean squared error of the estimates are calculated from 3000 independent replications.

Table 1.3: Finite-sample properties of QML, EF, and FOSLS under Gamma(2,1)

T	Estimator	$\bar{\hat{\theta}}_0$	$RMSE_{\hat{\theta}_0}$	$\bar{\hat{\phi}}_0$	$RMSE_{\hat{\phi}_0}$	$\bar{\hat{\theta}}_0$	$RMSE_{\hat{\theta}_0}$	$\bar{\hat{\phi}}_0$	$RMSE_{\hat{\phi}_0}$
		Panel (a): $\theta_0 = \phi_0 = 0.2$				Panel (b): $\theta_0 = 0.2$, $\phi_0 = 0.6$			
60	QML	0.21	0.149	0.30	0.213	0.19	0.152	0.58	0.163
	EF	0.21	0.120	0.26	0.162	0.19	0.123	0.57	0.141
	FOSLS	0.21	0.119	0.27	0.172	0.19	0.121	0.59	0.136
100	QML	0.20	0.116	0.25	0.169	0.20	0.121	0.58	0.134
	EF	0.21	0.090	0.23	0.133	0.20	0.094	0.58	0.108
	FOSLS	0.20	0.091	0.23	0.138	0.20	0.094	0.59	0.108
1000	QML	0.20	0.038	0.20	0.063	0.20	0.040	0.60	0.037
	EF	0.20	0.029	0.20	0.051	0.20	0.028	0.60	0.030
	FOSLS	0.20	0.029	0.20	0.052	0.20	0.028	0.60	0.030
		Panel (c): $\theta_0 = 0.8$, $\phi_0 = 0.2$				Panel (d): $\theta_0 = 0.8$, $\phi_0 = 0.6$			
60	QML	0.77	0.098	0.29	0.215	0.77	0.097	0.59	0.161
	EF	0.79	0.070	0.25	0.165	0.78	0.079	0.57	0.151
	FOSLS	0.78	0.073	0.27	0.178	0.78	0.077	0.59	0.139
100	QML	0.78	0.074	0.25	0.172	0.78	0.071	0.58	0.128
	EF	0.79	0.053	0.23	0.134	0.79	0.053	0.58	0.112
	FOSLS	0.79	0.054	0.23	0.140	0.79	0.054	0.59	0.106
1000	QML	0.80	0.021	0.20	0.062	0.80	0.021	0.60	0.038
	EF	0.80	0.016	0.20	0.049	0.80	0.014	0.60	0.030
	FOSLS	0.80	0.016	0.19	0.050	0.80	0.015	0.60	0.030

Model: $y_t = \theta_0 y_{t-1} + \epsilon_t$, $\sigma_t^2 = 1 - \phi_0 + \phi_0 \epsilon_{t-1}^2$, and ϵ_t/σ_t are drawn independently from the standardized Gamma distribution with shape parameter equals to two. The sample mean and root mean squared error of the estimates are calculated from 3000 independent replications.

almost everywhere. There is no significant differences between the FOSLS and the EF estimators with respect to the RMSE or the bias.

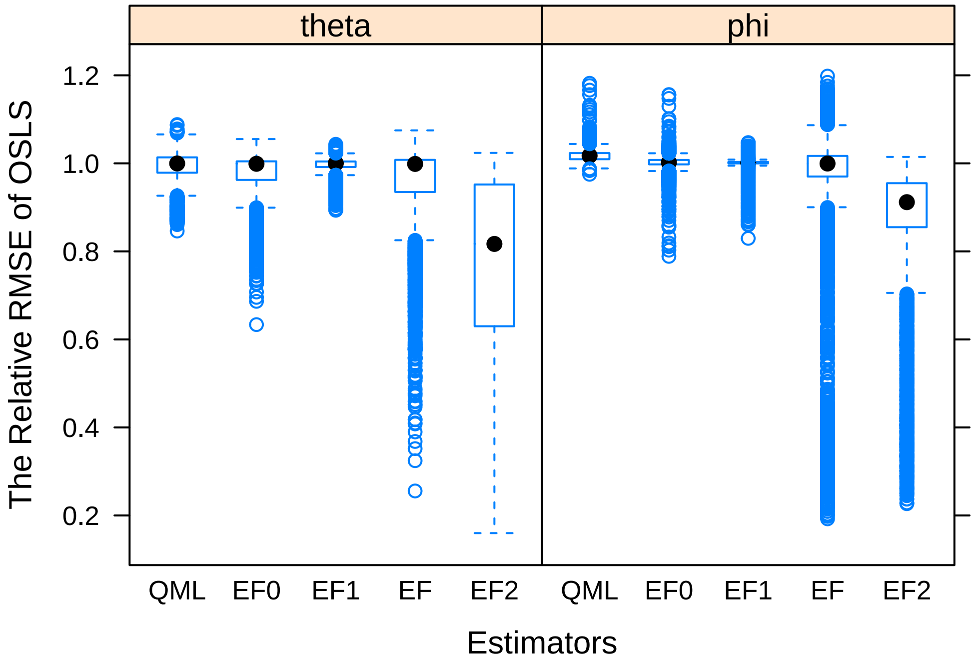
Figure 1.2-(a,b) (page 30) give an overall picture about the relative RMSE of the FOSLS compared to QML and four variants of the EF estimator, namely EF, EF0, EF1, EF2, as defined right after the graph.

Since ε_t are i.i.d., we can rewrite equation (1.12) in this form to produce the four EF variants,

$$\sum_{t=1}^{T} \boldsymbol{B}_t'(\boldsymbol{\gamma}_1)\boldsymbol{\Omega}_t^{-1}(\boldsymbol{\gamma}_2, \mu_3, \mu_4) \begin{pmatrix} \epsilon_t(\boldsymbol{\gamma}) \\ \epsilon_t^2(\boldsymbol{\gamma}) - \sigma_t^2(\boldsymbol{\gamma}) \end{pmatrix} = \mathbf{0}. \tag{1.13}$$

Then EF0 is obtained by taking $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2 = \hat{\boldsymbol{\gamma}}_T^Q$, and replacing the parameters μ_3, μ_4 with $1/T \sum_{t=1}^{T} \varepsilon_t^3(\hat{\boldsymbol{\gamma}}_T^Q)$, and $1/T \sum_{t=1}^{T} \varepsilon_t^4(\hat{\boldsymbol{\gamma}}_T^Q)$ respectively. EF1 is the same as EF0 except that $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}$, EF is the same as EF0 except that $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2 = \boldsymbol{\gamma}$, and EF2 is obtained by letting $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2 = \boldsymbol{\gamma}$ and replacing μ_3, μ_4 with $1/T \sum_{t=1}^{T} \varepsilon_t^3(\boldsymbol{\gamma})$, and $1/T \sum_{t=1}^{T} \varepsilon_t^4(\boldsymbol{\gamma})$ respectively. The four variants are consistent and sharing the same asymptotic variance. Figure 1.2-a is produced out from 4050 simulations with 3000 independent replications for each where the innovations are generated from standardized Student-t with 5, 6, 7, 8, 9 df, and T varies on 30, 40, 50, 60, 70, 80, 90, 100, 500, 1000. The RMSE of the QML, EF, EF0, EF1, EF2, and FOSLS are calculated on parameters grid $\{(0.1, 0.1)$, $(0.1, 0.2)$, $\ldots, (0.1, 0.9)$, $(0.2, 0.1), \ldots, (0.9, 0.9)\}$. In a similar way 4790 simulations are used to produce Figure 1.2-b except that the innovations are generated from standardized Gamma distribution with shape parameter 2, 3, 4, 5, 6, 7. Figure 1.2-a shows that the QML is a reasonable choice in case of symmetric distributions. However, it is not good choice in case of skewed

(a) $\varepsilon_t \sim$ Student-t with df (5, 6, 7, 8, 9)



(b) $\varepsilon_t \sim$ Gamma with shape parameters (2, 3, 4, 5, 6, 7)

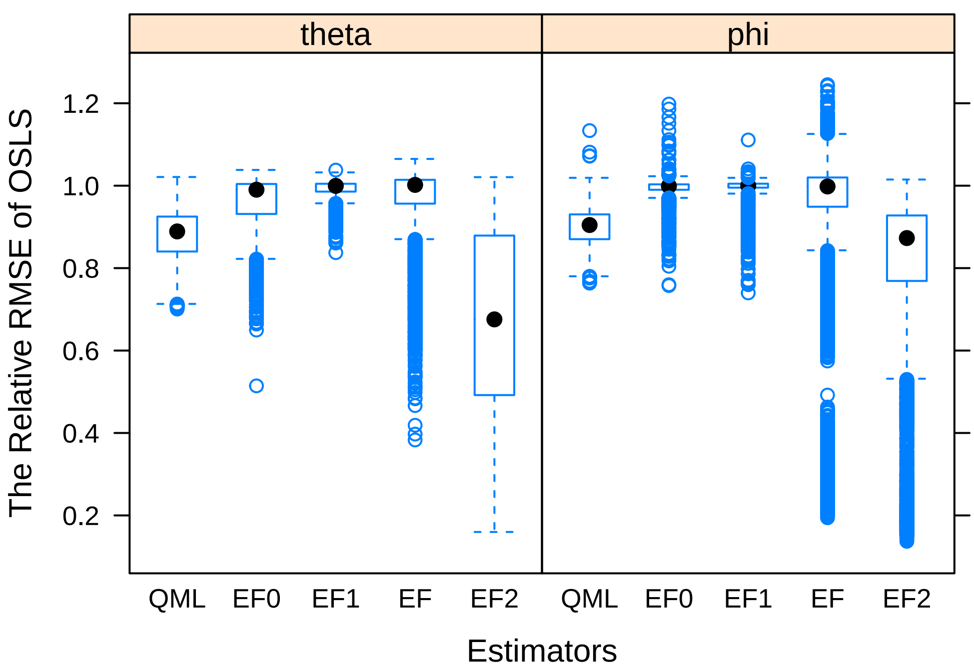


Figure 1.2: Relative RMSE of FOSLS compared to QML, and variants of EF

distributions as shown in Figure 1.2-b. It is also clear from Figure 1.2-(a,b) that the FOSLS is very good competitor to the EF estimator. Replacing the nuisance parameters with highly nonlinear generic functions of the estimated parameter makes the performance getting worse as seen with EF2. So, it is recommended to use two-step EF estimator such as the EF or EF1.

In order to understand how the values of (θ_0, ϕ_0), shape parameter, and T explain the relative RMSE of FOSLS compared to QML, two regression equations are fitted with RRMSE of FOSLS as a dependent variable. We use the results of the 4790 simulations which produce Figure 1.2-b. As shown from the summarized output in Table 1.4, the shape parameter has a significant positive effect on both RRMSE($\hat{\theta}^o_T$) and RRMSE($\hat{\phi}^o_T$). This is consistent with the fact that as the shape parameter gets larger the the gamma distribution becomes closer to the Gaussian distribution. This also can be seen from Table 1.1 in page 23. Moreover, an increase in the series length is associated with a significant decrease in both RRMSE($\hat{\theta}^o_T$) and RRMSE($\hat{\phi}^o_T$). This indicates that the over-performance of the FOSLS in large samples is more evident than in small samples if the innovations distribution is skewed. Last but not least, the performance of the $\hat{\theta}^Q_T$ gets better quickly as the value of θ_0 gets larger. This can be seen from the significant negative sign of θ_0 in the equation of RRMSE($\hat{\theta}^o_T$).

Table 1.4: The Effect of the shape, series length and the parameters values on the RRMSE of FOSLS compared to QML under Gamma distribution

	Coefficients						
	Constant	Shape	T	θ_0	ϕ_0	R^2	Error DF
$RRMSE(\hat{\theta}^o_T)$	0.76546	0.02887	-0.00007	-0.02783	0.02525	0.78692	4785
$RRMSE(\hat{\phi}^o_T)$	0.79952	0.02078	-0.00003	0.01013	0.00962	0.70135	4785

All the coefficients are significant at 0.0001 level of significance.

In addition to the Monte Carlo results which support the proposed estimator, we further investigate the practical usefulness of our estimator with a real data example in the next section.

1.6 Empirical Example: The U.K. Inflation

In this section we apply our method to the U.K. inflation data studied by Engle (1982) (see also Enders, 2010). In particular, Engle (1982) studied the wage/price spiral for the U.K. over the period 1958Q2-1977Q2 and proposed an AR model with ARCH error. Let p_t denote the log of the U.K. consumer price index and w_t denote the log of the index of the nominal wage rates. Then the rate of inflation is $y_t = p_t - p_{t-1}$ and the real wage is $r_t = w_t - p_t$. Engle (1982) fitted the following conditional mean equation of the U.K. inflation rate using the method of OLS under the assumption of homoeosdacticity where the standard errors are in parentheses

$$\begin{aligned} y_t = \; & \underset{(0.006)}{0.0257} + \underset{(0.103)}{0.334}\, y_{t-1} + \underset{(0.110)}{0.408}\, y_{t-4} - \underset{(0.114)}{0.404}\, y_{t-5} + \underset{(0.014)}{0.0559}\, r_{t-1} + \epsilon_t, \\ \hat{\sigma}_t^2 = \; & 8.9 \times 10^{-5} \end{aligned} \tag{1.14}$$

Since the Lagrange multiplier test for ARCH(1) disturbances of the model in (1.14) was not significant, but test for ARCH(4) process was significant, Engle specified the following conditional variance equation

$$\sigma_t^2 = \phi_0 + \phi_1(0.4\,\epsilon_{t-1}^2 + 0.3\,\epsilon_{t-2}^2 + 0.2\,\epsilon_{t-3}^2 + 0.1\,\epsilon_{t-4}^2) \tag{1.15}$$

where the two-parameter variance function with declining set of weights on the disturbances was chosen to ensure the nonnegativity and stationarity constraints.

Further, Engle used the method of ML to fit equations (1.14) and (1.15) jointly using simple iterative schema. All coefficients (except the first lag in the inflation rate) were significant at level of significance 0.05.

Since we could not obtain the wage rates before 1963, we use the data from 1963Q1 through 1982Q1 to compensate for the 19 missing quarters. The data are obtained form OECD website `http://dx.doi.org/10.1787/data-00052-en` (see OECD, "Main Economic Indicators - complete database", Main Economic Indicators).

We first computed OLS estimation on equation (1.14) to see if there is structural change or major difference compared to the data used by Engle (1982), which yields

$$\begin{array}{llllll} y_t = & \underset{(0.0165)}{0.059} & + \underset{(0.1067)}{0.3822}\, y_{t-1} & + \underset{(0.1108)}{0.3666}\, y_{t-4} & - \underset{(0.1153)}{0.3383}\, y_{t-5} & + \underset{(0.0193)}{0.0628}\, r_{t-1} + \epsilon_t, \\ \hat{\sigma}_t^2 = 13.3 \times 10^{-5} & & & & & \end{array} \tag{1.16}$$

Accordingly, there is no structural change by replacing the 19 quarters, which means that our results are to some extent comparable with those in (Engle, 1982).

Next we fit a mean equation with five lags of inflation using the method of OLS,

$$y_t = \theta_0 + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \theta_3 y_{t-3} + \theta_4 y_{t-4} + \theta_5 y_{t-5} + \theta_6 r_{t-1} + \epsilon_t \tag{1.17}$$

The results are shown in the first part of Table 1.5 (next page) under model-I (OLS), where the White's correction for the standard errors are reported in parentheses. The Ljung-Box statistic Q is calculated for $\hat{\varepsilon}_t$ (denoted by Q1) and $\hat{\varepsilon}_t^2$ (denoted by Q2) at lags 5, 10, 15, and 20. They (not listed in the table) are all insignificant at 0.1 level of significance except for Q2(5), which agrees with Engle's point of including four

Table 1.5: The models fitted to the price inflation using British data from 1963Q1 through 1982Q1.

	Model-I			Model-II	
Coef.	OLS	QML	FOSLS	QML	FOSLS
			Conditional Mean Equation		
θ_0	0.071^a	0.079^a	0.049^a	0.067^a	0.063^a
	(0.016)	(0.015)	(0.014)	(0.015)	(0.013)
θ_1	0.417^a	0.339^a	0.280^a	0.323^a	0.257^a
	(0.141)	(0.1)	(0.093)	(0.096)	(0.086)
θ_2	-0.039	0.004	0.050	–	–
	(0.095)	(0.086)	(0.081)		
θ_3	-0.180	-0.235^a	-0.158	–	–
	(0.148)	(0.101)	(0.094)		
θ_4	0.436^a	0.481^a	0.563^a	0.328^a	0.339^a
	(0.184)	(0.108)	(0.101)	(0.116)	(0.104)
θ_5	-0.350^a	-0.310^a	-0.294^a	-0.246^a	-0.234^a
	(0.101)	(0.094)	(0.088)	(0.094)	(0.084)
θ_6	0.076^a	0.086^a	0.051^a	0.073^a	0.067^a
	(0.018)	(0.017)	(0.016)	(0.017)	(0.015)
			Conditional Variance Equation		
ϕ_0	0.0001	0.000^a	0.000	0.000^a	0.000
		(0.000)	(0.000)	(0.000)	(0.000)
ϕ_1	0.1064	0.093	0.021	–	–
		(0.134)	(0.126)		
ϕ_2	0.0000	0.000	0.000	–	–
		(0.077)	(0.072)		
ϕ_3	0.0806	0.100	0.102	–	–
		(0.146)	(0.137)		
ϕ_4	0.3364	0.389	0.479^a	0.553	0.556^a
		(0.252)	(0.235)	(0.308)	(0.281)
			Diagnostic Statistics of the Standardized Innovations		
Q1(5)	0.9(0.97)	1.0(0.96)	2.1(0.83)	2.0(0.85)	1.9(0.86)
Q1(10)	4.4(0.93)	7.4(0.68)	6.4(0.78)	8.4(0.59)	8.5(0.58)
Q1(15)	12.8(0.62)	16.1(0.38)	14.1(0.52)	19.6(0.19)	18.7(0.23)
Q2(5)	3.7(0.59)	4.4(0.50)	0.4(0.99)	6.5(0.26)	6.4(0.27)
Q2(10)	5.7(0.84)	6.9(0.74)	1.3(0.99)	8.2(0.61)	8.1(0.62)
Q2(15)	9.2(0.87)	9.8(0.83)	2.7(0.99)	9.1(0.87)	9.4(0.85)
Skew.	0.78	0.61	1.67	0.73	0.99
Kurt.	4.07	3.49	8.48	3.73	4.27
JB	11.0^a	5.35	115^a	8.12^a	16.93^a

$Q1(n)$ denotes the Ljung-Box statistic for the standardized innovations up to lag n. $Q2(n)$ denotes the Ljung-Box statistic for the squared standardized innovations. The p-values of the test are reported in parentheses. Model-I is given by equations (1.17,1.18) and Model-II is given by equation (1.19, 1.20). JB is standard Jarque-Bera test, Skew. and Kur. are skewness and kurtosis values respectively, ^{a} statistically significant at 5% level. The standard errors of the coefficients in the mean and variance equations are reported in parentheses.

lags in the variance equation. Further, from this regression the squared residuals are used to fit an auxillary five-parameter ARCH(4) model for the conditional variance

$$\sigma_t^2 = \phi_0 + \phi_1\epsilon_{t-1}^2 + \phi_2\epsilon_{t-2}^2 + \phi_3\epsilon_{t-3}^2 + \phi_4\epsilon_{t-4}^2 \tag{1.18}$$

The results are shown in the second part of Table 1.5 under title Model-I (OLS). Again, the ARCH(4) model is confirmed by the LM test at significance level 0.05. We report only the OLS estimates of the variance function without standard errors because those estimates are only used as starting values to compute the QML. From the Q1 and Q2 statistics for the standaridized innovations obtained from model-I (OLS) (in the third part of Table 1.5), we can see that the mean and variance equations are fairly well specified since none of these diagnostics is significant at level 0.1. Therefore we fit model-I again using the QML approach which is more efficient than the two step OLS procedure. However, although the diagnostics of the standardized innovations obtained from model-I (QML) do not show serial correlation of first or second order, all coefficients in the variance function are insignificant except for the constant term. This contradicts with the ARCH(4) that we found before to be correctly specified. However, this can be explained by the lack of efficiency in the QML due to the moderate level of skewness in the corresponding residuals.

On the other hand, our FOSLS estimation yields significant fourth lag in the variance function in addition to the correct specification as seen by Q1 and Q2. Accordingly the model fitted by FOSLS is used in stepwise regression algorithm to obtain a reduced model (model-II in Table 1.5)

$$y_t = \theta_0 + \theta_1 y_{t-1} + \theta_4 y_{t-4} + \theta_5 y_{t-5} + \theta_6 r_{t-1} + \epsilon_t \tag{1.19}$$

$$\sigma_t^2 = \phi_0 + \phi_4 \epsilon_{t-4}^2 \tag{1.20}$$

Note that while the mean equation is identical to that in (Engle, 1982), only the fourth lag is significant in the variance equation. The above ARCH structure can only be detected by using the full model that is more flexible than the two-parameter variance equation in (Engle, 1982). Moreover, the more efficient FOSLS estimation yields the ARCH(4) structure, while the QML would conclude with a misspecified homoscedastic model.

1.7 Summary

Although ARCH-type models have been intensively studied for decades, most of theory is based on ARMA specification for the conditional mean of the underlying process. However, it is well-known that many economics and financial time series are nonlinear and/or nonstationary in mean. In this chapter we proposed a flexible and general model with dynamic structure in the mean and variance functions. The conditional mean of the process takes general form which covers the case of stationary linear as well as time variant nonlinear function. The conditional variance process is a standard ARCH(p). We generalized the SLS approach to this model based on the first two conditional moments of the underlying process. The root-T consistency is established under fairly general mixing process assumption. It has been demonstrated that the OSLS estimator is superior to the commonly used QMLE, and the efficiency gain is significant when the underlying distribution is asymmetric. Although the proposed optimal estimator is asymptotically as efficient as the optimal estimating function estimator based on the first two moments, our large scale simulation studies showed that it behaves better in finite sample situation. The chapter ends up with an empirical example in which we used popular data set to highlight the practical usefulness of the OSLS gain of efficiency over the QMLE.

Chapter 2

SLS Estimation in Linear Dynamic Panel Data Model

Dynamic panel data models have been intensively studied in econometrics (e.g. Chamberlain, 1984; Arellano and Honoré, 2001; Arellano, 2003; Hsiao, 2003; Baltagi, 2008). The literature show two main approaches which are being used to control for the unobserved subject effect in the additive dynamic panel data models. The first one is the fixed effects (FE) approach which does not assume any statistical model for relationship between the subject effect and the observed regressors. It uses some suitable linear transformation such as the first differencing or the forward orthogonal deviations to eliminate the unobserved effect term from the regression equation (see Arellano and Bover, 1995). Some commonly used GMM estimators such as the first-differenced GMM, and the system GMM estimators belong to this category (see Arellano and Bond, 1991; Blundell and Bond, 1998, respectively).

The second approach is the random effects (RE) which on the contrary to the FE approach postulates some assumptions on the relationship between the unobserved subject effect and the observed regressors. These assumptions are mainly about the conditional distribution of the unobserved effect given the observed regressors. Some commonly used pseudo-ML estimators such as the marginal PML, the conditional PML, and the random effects PML estimators belong to this category (see Arellano, 2003, ch. 6). The merits and demerits of these two approaches were discussed thoroughly by Hsiao (2011). Generally speaking, the RE approach uses both within-group and between group information to produce more efficient estimators, while the FE approach yields more robust estimators which are consistent, asymptotically normal, and computationally tractable under fairly general conditions. Some authors try to compromise between efficiency and robustness by using the RE approach with minimal assumptions on the conditional moments of the unobserved subject effect given the observed regressors. For example, Blundell and Bond (1998) studied the conditional generalized least squares (CGLS) estimator which assumes a simple linear relationship between the random effect and the initial observation.

In this chapter, we use the SLS approach to compromise between the RE and FE approaches in estimating a general linear dynamic model. The SLS estimator is defined in terms of the first two conditional moments of the response variable given its initial observation and the covariates. The data generating process can be either stationary or nonstationary. This approach requires only the specification of the first two conditional moments of the unobserved subject effect given the process initial value and the covariates, and does not require any other initial conditions or distributional assumptions. We derive the asymptotic properties of the SLS estimator when the cross section size; N, is large and the time series length; T, is fixed. Then, we show that the proposed estimator reaches a semiparametric efficiency bound in the sense of Chamberlain (1987). The correct specification of the first two conditional moments (mentioned above) is crucial assumption for the developed theory and any empirical application of the SLS. In order to double check such assumption we introduced for the first time a new specification test for the SLS estimator. Moreover, we conduct Monte Carlo simulations to investigate the finite sample performance of different variants of SLS estimator and compare them with other GMM and other likelihood based estimators. Finally, we used a real data set to show practical merits of using the SLS estimator compared to other commonly used estimators.

This chapter is organized into six sections. In the first section we introduce the model in details, and in the second section we define the SLS estimator and give its asymptotic properties with the regularity conditions. In the third section we discuss the asymptotic efficiency of the SLS estimator and some related computational matters including the newly proposed specification test. In the fourth section we present the simulation results and discussion. The fifth section is devoted for empirical example on the determinants of U.S. airfares. The last section gives a summary of the chapter. Related mathematical proofs, output tables, and graphs are all included in the appendices.

2.1 Model Specification and SLS Estimation

Let $(\boldsymbol{y}_i', \boldsymbol{x}_i', \eta_i)$, $i = 1, 2, \ldots, N$ be independent and identically distributed random vectors where $\boldsymbol{y}_i' = (y_{i0}, y_{i1}, \ldots, y_{iT})$ is the measurements of the response variable y taken for the ith subject over $T+1$ time periods, $\boldsymbol{x}_i' = (\boldsymbol{x}_{i1}', \boldsymbol{x}_{i2}', \ldots, \boldsymbol{x}_{iT}')$ is the corresponding measurements of p strictly exogenous regressors taken for the ith subject over T time periods such that

$$y_{it} = \alpha_0 y_{i(t-1)} + \boldsymbol{\beta}_0' \boldsymbol{x}_{it} + \eta_i + \varepsilon_{it}, \quad t = 1, 2, \ldots, T, \tag{2.1}$$

η_i is unobserved subject effect, the disturbance term ε_{it} satisfies $E(\varepsilon_{it}|\eta_i, y_{i0}, \boldsymbol{x}_i) = 0$, and $E(\varepsilon_{it}\varepsilon_{is}|\eta_i, y_{i0}, \boldsymbol{x}_i) = \sigma_0^2$ if $s = t$, zero otherwise. These assumptions are weaker than the frequently used *sequential moments* setup (see Arellano, 2003, ch. 6). In addition, we assume that the conditional moments $E(\eta_i^j|y_{i0}, \boldsymbol{x}_i) = f_j(y_{i0}, \boldsymbol{x}_i, \boldsymbol{\theta}_0)$, $j = 1, 2$ are known up to an ℓ-dimensional unknown parameter vector $\boldsymbol{\theta}_0$. Again, this later assumption is more general than the unrestricted initial conditions which are used by Blundell and Bond (1998) and Alvarez and Arellano (2003) to derive the conditional GLS (CGLS) and the random effects ML (RML) respectively. It is worthwhile to mention here that our semi-parametric assumption on the unobserved heterogeneity is not restrictive as might be thought. The functional form of $f_j(y_{i0}, \boldsymbol{x}_i, \boldsymbol{\theta}_0)$, $j = 1, 2$ can be specified naturally based on the residuals produced from equation (2.1). More important, any suggested specification can be tested using our developed test of specification.

Let $\boldsymbol{\gamma}_0 = (\alpha_0, \sigma_0^2, \boldsymbol{\beta}_0', \boldsymbol{\theta}_0')'$, and $\Gamma \subset I\!\!R^{p+\ell+2}$ be the corresponding parameter space. By backward substitution in equation (2.1), we obtain the reduced form equation

$$y_{it} = \alpha_0^t y_{i0} + a_t(\alpha_0)\eta_i + \boldsymbol{\beta}_0' \tilde{\boldsymbol{x}}_{it}(\alpha_0) + \tilde{\varepsilon}_{it}(\alpha_0), \quad t = 1, 2, ..., T, \tag{2.2}$$

where $a_t(\alpha) = \sum_{r=0}^{t-1} \alpha^r$, $\tilde{\boldsymbol{x}}_{it}(\alpha) = \sum_{r=0}^{t-1} \alpha^r \boldsymbol{x}_{i(t-r)}$, and $\tilde{\varepsilon}_{it}(\alpha_0) = \sum_{r=0}^{t-1} \alpha_0^r \varepsilon_{i(t-r)}$.

Under this model, the first two conditional moments of y_{it} given the initial observation y_{i0} and the exogenous measurements $\boldsymbol{x}_i$ are given by

$$\begin{aligned}
\mu_{it}(\boldsymbol{\gamma}_0) &= E(y_{it}|y_{i0}, \boldsymbol{x}_i) = \alpha_0^t y_{i0} + \boldsymbol{\beta}_0' \tilde{\boldsymbol{x}}_{it}(\alpha_0) + a_t(\alpha_0) f_1(y_{i0}, \boldsymbol{x}_i, \boldsymbol{\theta}_0), & (2.3)\\
\nu_{its}(\boldsymbol{\gamma}_0) &= E(y_{it} y_{is}|y_{i0}, \boldsymbol{x}_i) = \alpha_0^{t+s} y_{i0}^2 + a_t(\alpha_0) a_s(\alpha_0) f_2(y_{i0}, \boldsymbol{x}_i, \boldsymbol{\theta}_0) &\\
&\quad + \boldsymbol{\beta}_0' \tilde{\boldsymbol{x}}_{it}(\alpha_0) \tilde{\boldsymbol{x}}_{is}'(\alpha_0) \boldsymbol{\beta}_0 + \sigma_0^2 c_{ts}(\alpha_0) + d_{ts}(\alpha_0) y_{i0} f_1(y_{i0}, \boldsymbol{x}_i, \boldsymbol{\theta}_0) &\\
&\quad + y_{i0} \boldsymbol{\beta}_0' \boldsymbol{w}_{its}(\alpha_0) + f_1(y_{i0}, \boldsymbol{x}_i, \boldsymbol{\theta}_0) \boldsymbol{\beta}_0' \boldsymbol{k}_{its}(\alpha_0), & (2.4)
\end{aligned}$$

where

$$\begin{aligned}
&c_{ts}(\alpha) = \alpha^{t-s} \textstyle\sum_{r=0}^{s-1} \alpha^{2r}, &&\boldsymbol{w}_{its}(\alpha) = \alpha^t \tilde{\boldsymbol{x}}_{is}(\alpha) + \alpha^s \tilde{\boldsymbol{x}}_{it}(\alpha),\\
&d_{ts}(\alpha) = \alpha^t a_s(\alpha) + \alpha^s a_t(\alpha), &&\boldsymbol{k}_{its}(\alpha) = a_t(\alpha) \tilde{\boldsymbol{x}}_{is}(\alpha) + a_s(\alpha) \tilde{\boldsymbol{x}}_{it}(\alpha), \quad \text{for } t \geq s.
\end{aligned}$$

Let

$$\boldsymbol{h}_i(\boldsymbol{\gamma}) = (y_{it} - \mu_{it}(\boldsymbol{\gamma}), 1 \leq t \leq T,\ y_{it} y_{is} - \nu_{its}(\boldsymbol{\gamma}), 1 \leq s \leq t \leq T)',$$

where in the second part of $\boldsymbol{h}_i(\boldsymbol{\gamma})$ the t subscript is changing after the s subscript. Then following Wang (2007) the SLS estimator for $\boldsymbol{\gamma}$ is obtained by

$$\hat{\boldsymbol{\gamma}}_N = \operatorname*{argmin}_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \frac{1}{N}\sum_{i=1}^{N} q_i(\boldsymbol{\gamma}), \tag{2.5}$$

where $q_i(\boldsymbol{\gamma}) = \boldsymbol{h}_i'(\boldsymbol{\gamma})\boldsymbol{W}_i\boldsymbol{h}_i(\boldsymbol{\gamma})$ and $\boldsymbol{W}_i$ is a nonnegative definite matrix whose elements are Borel measurable functions of $(y_{i0}, \boldsymbol{x}_i)$.

The asymptotic properties of $\hat{\boldsymbol{\gamma}}_N$ are studied in the next section.

2.2 Asymptotic Properties of the SLS Estimator

For the existence, measurability, consistency and asymptotic normality of the $\hat{\boldsymbol{\gamma}}_N$, we make the following assumptions, where $\|\cdot\|$ denotes the Euclidean norm.

Assumption 13 *The parameter space $\boldsymbol{\Gamma}$ is a compact subset of $\mathbb{R}^{p+\ell+2}$.*

Assumption 14 $f_j(y_{i0}, \boldsymbol{x}_i, \boldsymbol{\theta}), j = 1, 2$ *are Borel measurable functions of* $(y_{i0}, \boldsymbol{x}_i)$ *for each $\boldsymbol{\theta}$ in the corresponding parameter space $\boldsymbol{\Theta}$, and are continuous functions of $\boldsymbol{\theta}$ with probability one. Furthermore, for all t*

$$E\,\|\boldsymbol{W}_1\|\left(y_{10}^4 + \sup_{\boldsymbol{\Theta}} f_2^2(y_{10}, \boldsymbol{x}_1, \boldsymbol{\theta}) + \|\boldsymbol{x}_{1t}\|^4 + \eta_1^4 + \varepsilon_{1t}^4 + 1\right) < \infty.$$

Assumption 15 $E[\boldsymbol{h}_1(\boldsymbol{\gamma}) - \boldsymbol{h}_1(\boldsymbol{\gamma}_0)]'\boldsymbol{W}_1[\boldsymbol{h}_1(\boldsymbol{\gamma}) - \boldsymbol{h}_1(\boldsymbol{\gamma}_0)] = 0$ *if and only if* $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$.

Assumption 16 *With probability one, $f_j(y_{i0}, \boldsymbol{x}_i, \boldsymbol{\theta}), j = 1, 2$ are twice continuously differentiable in int($\boldsymbol{\Theta}$). Furthermore, for $j = 1, 2$,*

$$E \left\| \boldsymbol{W}_1 \right\| (y_{10}^2 + \left\| \boldsymbol{x}_1 \right\|^2 + 1)^{2-j} \sup_{\mathcal{N}(\boldsymbol{\theta}_0)} \left(\left\| \frac{\partial f_j(y_{10}, \boldsymbol{x}_1, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|^2 + \left\| \frac{\partial^2 f_j(y_{10}, \boldsymbol{x}_1, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\|^2 \right) < \infty,$$

where $\mathcal{N}(\boldsymbol{\theta}_0) \subset int(\boldsymbol{\Theta})$ is a closed neighbourhood of $\boldsymbol{\theta}_0$.

Assumption 17 *The matrix*

$$\boldsymbol{A} = E \left\{ \frac{\partial \boldsymbol{h}_1'(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}} \boldsymbol{W}_1 \frac{\partial \boldsymbol{h}_1(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}'} \right\} \tag{2.6}$$

is nonsingular.

Assumption 18 *It holds*

$$E \left\| \boldsymbol{W}_1 \right\|^2 \left(y_{10}^8 + \left\| \boldsymbol{x}_1 \right\|^8 + f_2^4(y_{10}, \boldsymbol{\theta}_0) + \eta_1^8 + \varepsilon_{1t}^8 \right.$$
$$\left. + \left(y_{10}^4 + \left\| \boldsymbol{x}_1 \right\|^4 + 1 \right) \left\| \frac{\partial f_1(y_{10}, \boldsymbol{x}_1, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\|^4 + \left\| \frac{\partial f_2(y_{10}, \boldsymbol{x}_1, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\|^4 + 1 \right) < \infty.$$

Then we have the following two propositions whose proofs are found in appendix A.

Proposition 3 *Under assumptions 13–15, $\hat{\boldsymbol{\gamma}}_N \xrightarrow{a.s.} \boldsymbol{\gamma}_0$ as $N \to \infty$ for fixed T.*

Proposition 4 *Under assumptions 13–18, $\sqrt{N}(\hat{\boldsymbol{\gamma}}_N - \boldsymbol{\gamma}_0) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1})$ as $N \to \infty$ and T is fixed, where*

$$\boldsymbol{B} = E \left\{ \frac{\partial \boldsymbol{h}_1'(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}} \boldsymbol{W}_1 \boldsymbol{h}_1(\boldsymbol{\gamma}_0) \boldsymbol{h}_1'(\boldsymbol{\gamma}_0) \boldsymbol{W}_1 \frac{\partial \boldsymbol{h}_1(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}'} \right\} \tag{2.7}$$

and $\boldsymbol{A}$ is given in assumption 17.

Note that assumptions 13-18 are standard regularity conditions in the M-estimation literature. In particular, assumption 15 is necessary and sufficient for parameters identification, while assumptions 14, 16, and 18 are sufficient but not necessary. Moreover, since these assumptions depend on the form of $f_j(y_{10}, \boldsymbol{x}_1, \boldsymbol{\theta}), j = 1, 2,$ and $\boldsymbol{W}_1$, they can be significantly simplified under extra specifications of model (2.1) as emphasized in the next section.

2.3 Asymptotically Optimal SLS

From equation (2.7) we see that the asymptotic covariance of $\hat{\boldsymbol{\gamma}}_N$ depends on the weight matrix $\boldsymbol{W}_i$. In this section we derive the asymptotically optimal (efficient) SLS estimator by choosing the optimal weight matrix that minimizes the asymptotic variance of $\sqrt{N}\boldsymbol{a}'(\hat{\boldsymbol{\gamma}}_N - \boldsymbol{\gamma}_0)$, for all $\boldsymbol{a} \in \mathbb{R}^{\ell+2}$.

Proposition 5 *Suppose $\boldsymbol{U}_1 = E\left\{\boldsymbol{h}_1(\boldsymbol{\gamma}_0)\boldsymbol{h}_1'(\boldsymbol{\gamma}_0) \middle| y_{10}, \boldsymbol{x}_1\right\}$ is nonsingular with probability one, and assumptions 14–18 are satisfied with $\boldsymbol{W}_1 = \boldsymbol{U}_1^{-1}$. Then the asymptotically optimal SLS estimator is obtained by taking $\boldsymbol{W}_i = \boldsymbol{U}_i^{-1}, i = 1, 2, ..., N$.*

The proof is straightforward by noting that $E\left(\boldsymbol{R}'\boldsymbol{R}\right) - E\left(\boldsymbol{R}'\boldsymbol{Q}\right)E^{-1}\left(\boldsymbol{Q}'\boldsymbol{Q}\right)E\left(\boldsymbol{Q}'\boldsymbol{R}\right)$ is nonnegative definite and it is zero matrix if $\boldsymbol{W}_1 = \boldsymbol{U}_1^{-1}$, where

$$\boldsymbol{R} = \boldsymbol{U}_1^{1/2}\boldsymbol{W}_1\frac{\partial \boldsymbol{h}_1(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}'}, \qquad \boldsymbol{Q} = \boldsymbol{U}_1^{-1/2}\frac{\partial \boldsymbol{h}_1(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}'}.$$

Since the optimal weight $\boldsymbol{U}_i^{-1}$ depends on $\boldsymbol{\gamma}_0$, $E(\varepsilon_{it}^j|y_{i0}, \boldsymbol{x}_i)$ and $E(\eta_i^j|y_{i0}, \boldsymbol{x}_i), j = 3, 4$, the optimal SLS estimator is not feasible. The corresponding feasible SLS estimator can be calculated by plugging in consistent estimators of these unknown quantities in $\boldsymbol{U}_i^{-1}$. This feasible optimal SLS (FOSLS) estimator can be regarded as a two-step M-estimator which is consistent under the following regularity condition

$$E \sup_{\boldsymbol{\Gamma}^*} \left\| \boldsymbol{U}_1^{-1}(\boldsymbol{\gamma}^*) \right\| \left(y_{10}^4 + \|\boldsymbol{x}_1\|^4 + \sup_{\boldsymbol{\Theta}} f_2^2(y_{10}, \boldsymbol{x}_1, \boldsymbol{\theta}) + \eta_1^4 + \varepsilon_{1t}^4 + 1 \right) < \infty, \quad (2.8)$$

where $\boldsymbol{\gamma}^*$ is a vector containing all the generic parameters appearing in $\boldsymbol{U}_1^{-1}$ including $\boldsymbol{\gamma}$ itself, and $\boldsymbol{\Gamma}^*$ is the corresponding compact parameter space. Moreover, we can use theorem 6.1 of Newey and McFadden (1994) to show that the FOSLS and the (infeasible) optimal SLS estimators have identical asymptotic distributions.

Although assumptions 14–18 and condition (2.8) look complicated, they could be greatly simplified by specifying the functional forms of $E(\eta_i^j|y_{i0}, \boldsymbol{x}_i), j = 1, 2, 3, 4$, and $E(\varepsilon_{it}^j|y_{i0}, \boldsymbol{x}_i), j = 3, 4$. To illustrate this point, we consider a common specification of model (2.1) which is often used to derive the conditional-type estimators such as the CGLS (Blundell and Bond, 1998) and the RML (Alvarez and Arellano, 2003). To simplify the notations and comparisons with other estimation approaches found in the literature, we consider model (2.1) without the regressor term $\boldsymbol{x}_{it}$ under the following two assumptions.

Assumption 19 $E(\eta_i|y_{i0}) = \theta_{01} + \theta_{02}y_{i0}$, *and* $V(\eta_i|y_{i0}) = \exp(\theta_{03})$.

Assumption 20 $E(\varepsilon_{it}^j|y_{i0}) = \mu_{j(\varepsilon)}$, *and* $E\left((\eta_i - \theta_{01} - \theta_{02}y_{i0})^j|y_{i0}\right) = \mu_{j(\eta)}, j = 3, 4$.

Under these two assumptions $\boldsymbol{U}_i^{-1}$ has a special structure so that the regularity assumptions 14, 16, 18 and condition (2.8) are implied by $E(y_{i0}^4) < \infty$. Moreover, in light of assumption 19 we can rewrite the identification assumption 15 as

$$E(y_{i0}^4) > \frac{\left[E(y_{i0}^2)\right]^3 + \left[E(y_{i0}^3)\right]^2 - 2E(y_{i0})E(y_{i0}^2)E(y_{i0}^3)}{V(y_{i0})}, \quad \text{and } T \geq 2.$$

To compute the FOSLS estimator, we suggest to use the RML approach to get a preliminary consistent estimates of $\boldsymbol{\gamma}_0$, $\mu_{j(\varepsilon)}$ and $\mu_{j(\eta)}, j = 3, 4$, then plug in those estimates in $\boldsymbol{U}_i^{-1}$. This does not affect the asymptotic properties as mentioned before.

We also suggest another version of the FOSLS, called FOSLS1, which may be more robust to any possible stochastic dependence between ε_{it} and η_i, and does not require initial estimates for $\mu_{j(\varepsilon)}$, $\mu_{j(\eta)}, j = 3, 4$. It is obtained by using the following weight matrix

$$\widehat{\boldsymbol{W}}_i = \boldsymbol{C}'\left(y_{i0}, \hat{\boldsymbol{\theta}}_N^0, \hat{\alpha}_N^0\right)\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{h}_i^*(\hat{\boldsymbol{\gamma}}_N^0)\boldsymbol{h}_i^{*'}(\hat{\boldsymbol{\gamma}}_N^0)\right)^{-}\boldsymbol{C}\left(y_{i0}, \hat{\boldsymbol{\theta}}_N^0, \hat{\alpha}_N^0\right), \qquad (2.9)$$

where $\hat{\boldsymbol{\gamma}}_N^0$ is the preliminary consistent estimator of $\boldsymbol{\gamma}_0$, and $C\left(y_{i0}, \boldsymbol{\theta}, \alpha\right)$ is a transformation matrix which maps $\boldsymbol{h}_i(\boldsymbol{\gamma})$ into

$$\boldsymbol{h}_i^*(\boldsymbol{\gamma}) = \left(u_{it}^*, 1 \leq t \leq T,\ u_{it}^* u_{is}^* - \nu_{its}^*(\sigma^2, \theta_3), 1 \leq s \leq t \leq T\right)',$$

where $u_{it}^* = y_{it} - \alpha y_{it-1} - f_1(y_{i0}, \boldsymbol{\theta})$ and $\nu_{its}^*(\sigma^2, \theta_3) = \exp(\theta_3) + \sigma^2 1_{\{s=t\}}$. It can be shown that under assumption 20, the FOSLS1 has the same asymptotic distribution of the infeasible OSLS estimator, however it may have different rate of convergence compared to the FOSLS.

A group of researchers have been searching for the semiparametric efficient estimators for the dynamic panel data model under different setups. Chamberlain (1992) derived the optimal instrumental variables for the first difference equation of model (2.1) under the sequential conditional moment restrictions. He showed that the GMM estimator based on these optimal instrumental variables attains the semiparametric efficiency bound of the model. Unfortunately, his optimal instrumental variables involve various conditional expectation functions which need to be imputed by nonparametric techniques. Hahn (1997) showed that the GMM estimator based on an increasing set of instruments as the sample size grows would achieve the semiparametric efficiency bound calculated by Chamberlain (1992). He also discussed the rate of growth of the number of instruments for the case of using the Fourier series and polynomial series as instruments. More recently, Park et al. (2007) used the geometric approach of Bickel et al. (1993) to construct an estimator which attains the semiparametric efficiency bound under the RE modeling approach. They assumed stationarity of the process, independence between the subject effect and the initial observation, normality distribution of the residual errors, and unknown distribution of the subject effect.

A natural question arises at this point. Does the optimal SLS estimator efficiently use the information inherent in the conditional moments $E\left\{\boldsymbol{h}_i(\boldsymbol{\gamma}_0)|y_{i0},\boldsymbol{x}_i\right\}=\mathbf{0}$?. To answer this question we use Chamberlain (1987) framework to derive the efficiency bound under $E\left\{\boldsymbol{h}_i(\boldsymbol{\gamma}_0)|y_{i0},\boldsymbol{x}_i\right\}=\mathbf{0}$. According to lemma 2 of Chamberlain (1987), the minimum bound of the asymptotic variance under these conditional moments is given by

$$
\begin{aligned}
\boldsymbol{F}_0^* &= E^{-1}\left\{E\left(\frac{\partial \boldsymbol{h}_i'(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}}\,|y_{i0},\boldsymbol{x}_i\right)E^{-1}\left(\boldsymbol{h}_i(\boldsymbol{\gamma}_0)\boldsymbol{h}_i'(\boldsymbol{\gamma}_0)|y_{i0},\boldsymbol{x}_i\right)E\left(\frac{\partial \boldsymbol{h}_i(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}'}|y_{i0},\boldsymbol{x}_i\right)\right\} \\
&= E^{-1}\left\{\frac{\partial \boldsymbol{h}_i'(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}}\boldsymbol{U}_i^{-1}\frac{\partial \boldsymbol{h}_i(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}'}\right\}. \qquad (2.10)
\end{aligned}
$$

According to proposition (5), the optimal SLS estimator attains Chamberlain's bound of variance, therefore it is a semiparametric optimal estimator in that sense.

We complete this discussion about the SLS efficiency by comparing the asymptotic variance of the optimal SLS estimator given in (2.10) with the asymptotic variance of the RML estimator given by equation (A.11) in the appendix A. Note that the RMLE is identical to the MLE conditional on the first observation when the error components are normally distributed. Theoretically the optimal SLSE is not less efficient than the RMLE as indicated by equations (A.11, A.12 and A.13) in the appendix A. Unfortunately it is difficult to investigate the gain of efficiency due to the optimal SLSE analytically, therefore we calculate the asymptotic variances of the two estimators and compare them numerically. The percentage gain of efficiency in estimating α_0 as a function of T and α_0 is calculated as shown in Figure 2.1, where z-axis represents the percentage reduction in the variance of RMLE(α_0) if we alternatively use the optimal SLSE(α_0). Many scenarios are considered wherein we change the data generating process (stationary or nonstationary), the distribution of the disturbance term, and the the distribution of the unobserved subject effect. Our observation indicates strongly that the asymptotic variance of the optimal SLSE(α_0) is strictly less than that of the RMLE(α_0) except for the case $\mu_{3(\varepsilon)} = 0$ and $\mu_{4(\varepsilon)} = 3\sigma_0^4$ (which is true if the disturbance term has a normal distribution), in

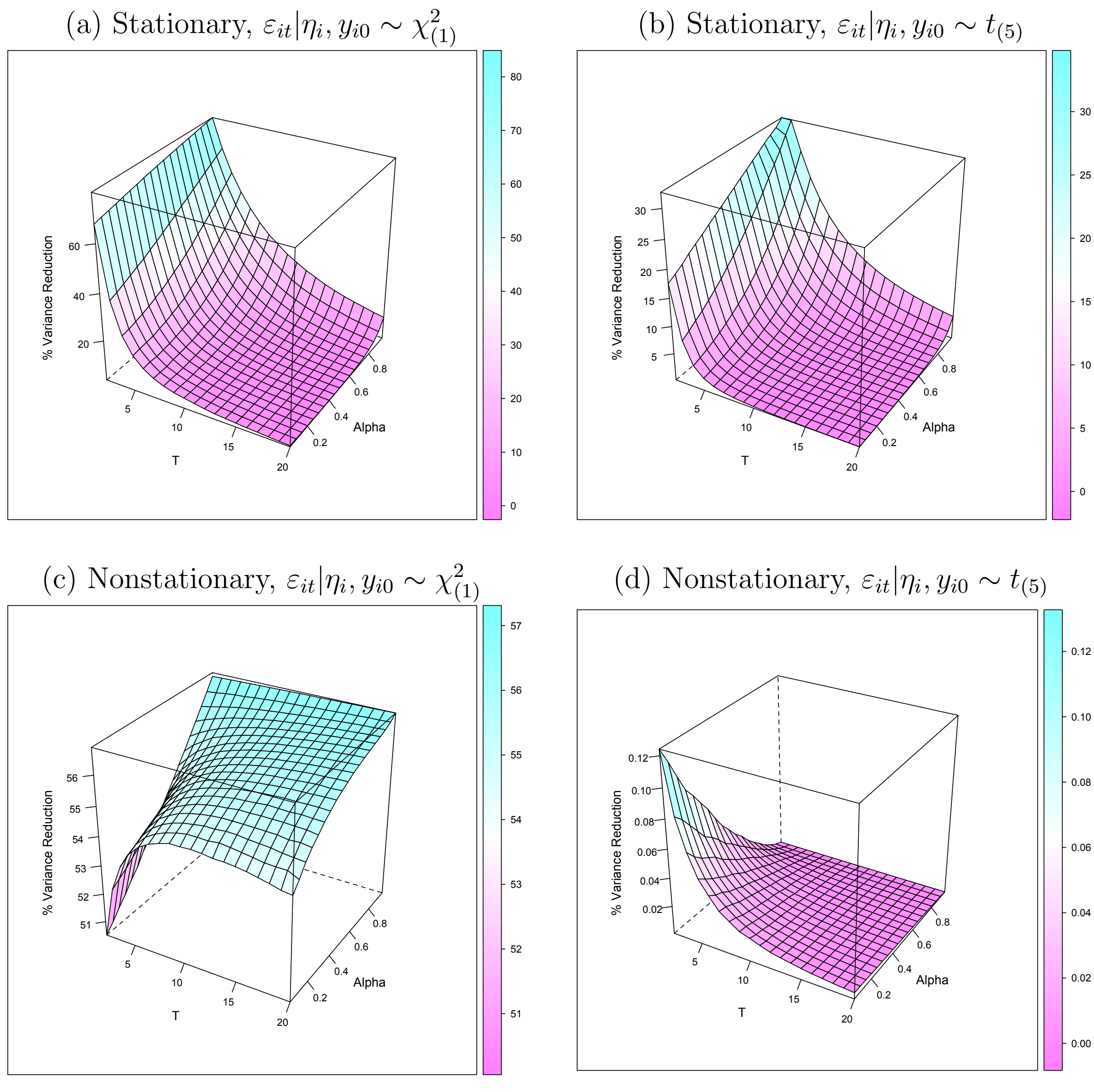


Figure 2.1: The reduction (%) in the variance of RML(α_0) gained by the OSLS

which case both estimators have the same asymptotic variances.

We conclude this section by introducing our proposed specification test named SW. It is designed to test for

$$H_0: \; E\{\boldsymbol{h}_i(\boldsymbol{\gamma}_0)|y_{i0}, \boldsymbol{x}_i\} = \boldsymbol{0} \quad vs. \quad H_a: \; E\{\boldsymbol{h}_i(\boldsymbol{\gamma}_0)|y_{i0}, \boldsymbol{x}_i\} \neq \boldsymbol{0},$$

where

$$\boldsymbol{h}_i(\boldsymbol{\gamma}) = (y_{it} - \mu_{it}(\boldsymbol{\gamma}), 1 \leq t \leq T, \; y_{it}y_{is} - \nu_{its}(\boldsymbol{\gamma}), 1 \leq s \leq t \leq T)',$$

and

$$\mu_{it}(\boldsymbol{\gamma}_0) = E(y_{it}|y_{i0}, \boldsymbol{x}_i), \quad \nu_{its}(\boldsymbol{\gamma}_0) = E(y_{it}y_{is}|y_{i0}, \boldsymbol{x}_i).$$

Define $\boldsymbol{h}_N(\boldsymbol{\gamma}) = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{h}_i(\boldsymbol{\gamma})$, then the test statistic SW is given by

$$\mathrm{SW} = N\, \boldsymbol{h}'_N(\hat{\boldsymbol{\gamma}}_N)\, \hat{\boldsymbol{G}}_N^{-1}\, \boldsymbol{h}_N(\hat{\boldsymbol{\gamma}}_N)$$

where $\hat{\boldsymbol{\gamma}}_N$ is defined by equation (2.5) and $\hat{\boldsymbol{G}}_N$ is given by

$$\hat{\boldsymbol{G}}_N = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{P}_i(\hat{\boldsymbol{\gamma}}_N)\boldsymbol{h}_i(\hat{\boldsymbol{\gamma}}_N)\boldsymbol{h}'_i(\hat{\boldsymbol{\gamma}}_N)\boldsymbol{P}'_i(\hat{\boldsymbol{\gamma}}_N),$$

where

$$\boldsymbol{P}_i(\hat{\boldsymbol{\gamma}}_N) = \boldsymbol{I} - \bar{\boldsymbol{D}}_N(\hat{\boldsymbol{\gamma}}_N)\boldsymbol{A}_N^{-1}\frac{\partial \boldsymbol{h}'_i(\hat{\boldsymbol{\gamma}}_N)}{\partial \boldsymbol{\gamma}}\boldsymbol{W}_i,$$

$$\boldsymbol{A}_N = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial \boldsymbol{h}'_i(\hat{\boldsymbol{\gamma}}_N)}{\partial \boldsymbol{\gamma}}\boldsymbol{W}_i\frac{\partial \boldsymbol{h}_i(\hat{\boldsymbol{\gamma}}_N)}{\partial \boldsymbol{\gamma}'},$$

and

$$\bar{\boldsymbol{D}}_N(\hat{\boldsymbol{\gamma}}_N) = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial \boldsymbol{h}_i(\hat{\boldsymbol{\gamma}}_N)}{\partial \boldsymbol{\gamma}'}$$

Under H_0, it can be shown that

$$\text{SW} \xrightarrow{d} \chi^2_{T(T+3)/2} \quad \text{as } N \to \infty \text{ and } T \text{ is fixed.}$$

The proof is straightforward by applying the mean value theorem for random functions on two stages. We note here that the SW test can be used to test a general specification of the first two conditional moments, even more general than the specification given by equations (2.3, 2.4). More interesting, the test doesn't postulate either using the optimal weight matrix in calculating $\hat{\boldsymbol{\gamma}}_N$ nor having initial estimates for $\mu_{j(\varepsilon)}$, $\mu_{j(\eta)}, j = 3, 4$. Moreover, the test can generalized and used in other platforms of SLS estimation. In the next section we examine the small sample properties of the SLS estimators.

2.4 Monte Carlo Simulations

In this section we carry out some Monte Carlo simulations to examine the finite sample performance of the two feasible OSLS estimators, namely FOSLS and FOSLS1. We also compare them with the marginal pseudo maximum likelihood (MPML) (Arellano, 2003), the linear first differenced GMM (Arellano and Bond, 1991), and the random effects pseudo maximum likelihood (RML) (Alvarez and Arellano, 2003) estimators. Following those authors we assume that the data are generated according to model (2.1) without covariates and after adding assumptions 19 and 20.

Specifically, $y_{i0} \overset{\text{i.i.d}}{\sim} \mathrm{N}\left(0, \frac{2}{(1-\alpha_0^2)(1-\alpha_0)}\right)$, $\eta_i|y_{i0} = \theta_{01} + \theta_{02} y_{i0} + \exp(\theta_{03}) F_1$, $\varepsilon_{it}|\eta_i, y_{i0} \overset{\text{i.i.d}}{=} \sigma_0^2 F_2$, $\theta_{01} = 0$, $\theta_{02} = c\left(\frac{1-\alpha_0^2}{2}\right)$, and $\theta_{03} = \log\left(\frac{1-\alpha_0}{2}\right)$. We consider the following scenarios.

Normal stationary process: $c = 1$, and $F_1,\ F_2 \sim \mathrm{N}(0, 1)$. Under this setup, the MPML estimator is the true MLE computed using all data including the initial observations.

Nonnormal stationary process: $c = 1$, and $F_1,\ F_2 \sim (\chi^2_{(1)} - 1)/\sqrt{2}$.

Normal nonstationary process: $c = 20$, and $F_1,\ F_2 \sim \mathrm{N}(0, 1)$. This choice of c makes the y process nonstationary with respect to the first and second moments. Under this setup the MPML estimator is inconsistent and the results demonstrate the consequence of misspecifying a nonstationary process.

Nonnormal nonstationary process:

(a) $c = 20$, $F_1 \sim \mathrm{N}(0, 1)$, and $F_2 \sim (\chi^2_{(1)} - 1)/\sqrt{2}$.

(b) $c = 20$, $F_1 \sim \mathrm{N}(0, 1)$, and $F_2 \sim \sqrt{3/5}\, t_{(5)}$.

(c) $c = 20$, $F_2 \sim \mathrm{N}(0, 1)$, and $F_1 \sim (\chi^2_{(1)} - 1)/\sqrt{2}$.

(d) $c = 20$, $F_2 \sim \mathrm{N}(0, 1)$, and $F_1 \sim \sqrt{3/5}\, t_{(5)}$.

In all scenarios the sample sizes are $N = 30$, 300 and $T = 5$, 10, 15, and the parameter values are $\alpha_0 = 0.2$, 0.5, 0.8. Five criteria were used to assess the finite sample performance of the given estimators, namely, the mean bias, median bias, root mean squared error, median absolute deviation (MAD), and the interquartile range.

All results are based on 1000 Monte Carlo replications. To allow for comparing our results with literature, some of the specifications are chosen to be very close to those in (Blundell and Bond, 1998; Alvarez and Arellano, 2003; Okui, 2009).

Tables 2.1–2.4 (next page) give the simulation results of the first three scenarios and the nonnormal nonstationary (a) scenario respectively. The output tables for the other scenarios are not produced to save space but their results are mentioned in the discussion below. In the tables we report the MADs and the medians of the estimators. The MADs are computed relatively to the FOSLS estimator which is taken as reference. Other criteria are not reported to save space.

Results and Discussion

Table 2.1 shows that the MAD of the FOSLS is decreasing in T and nondecreasing in α_0. The fact that the FOSLS and the RML have equal asymptotic variance under the normality of the error components (see A.13) appears clearly for $\alpha_0 = 0.2, 0.5$. Larger values of N are required to see this fact for larger α_0. The MPML has the smallest MAD almost everywhere. This is consistent with theory because the MPML is the most efficient estimator under normality. The gap between the MPML and the FOSLS is getting smaller as α_0 decreases or T increases. The linear first-differenced GMM is generally inferior, and the problem of weak instruments appears clearly for large α_0, which is consistent with (Blundell and Bond, 1998). The FOSLS1 is not reliable for small N. This is because the robust estimate of the optimal weight matrix is not well behaved for $N < T(T+3)/2$ (see 2.9). Moreover, the relative MADs of the FOSLS1 for $N = 300$ show that the estimator has a slower convergence

Table 2.1: Relative MAD and Median of the Estimators (Normal Stationary)

	$\alpha_0 = 0.2$					$\alpha_0 = 0.5$					$\alpha_0 = 0.8$				
	FOSLS	FOSLS1	RML	GMM	MPML	FOSLS	FOSLS1	RML	GMM	MPML	FOSLS	FOSLS1	RML	GMM	MPML
N								$T = 5$							
30	**0.112**	**1.02**	**1.01**	**1.59**	**1.01**	**0.137**	**1.01**	**0.97**	**1.77**	**0.84**	**0.130**	**1.23**	**1.34**	**4.72**	**0.90**
	0.170	0.17	0.16	0.12	0.19	0.490	0.47	0.46	0.37	0.50	0.770	0.72	0.69	0.39	0.80
300	**0.038**	**0.96**	**1.00**	**1.31**	**0.94**	**0.044**	**1.00**	**1.01**	**1.65**	**0.77**	**0.073**	**0.99**	**0.93**	**2.41**	**0.54**
	0.200	0.20	0.20	0.19	0.20	0.500	0.50	0.49	0.48	0.50	0.800	0.80	0.78	0.71	0.80
								$T = 10$							
30	**0.063**	**1.23**	**0.98**	**1.51**	**1.01**	**0.070**	**1.36**	**1.03**	**2.17**	**0.96**	**0.084**	**1.18**	**1.02**	**3.62**	**0.75**
	0.190	0.18	0.19	0.15	0.20	0.490	0.49	0.48	0.40	0.50	0.800	0.78	0.76	0.60	0.80
300	**0.021**	**0.97**	**1.00**	**1.27**	**0.96**	**0.022**	**1.09**	**0.99**	**1.48**	**0.92**	**0.029**	**1.17**	**0.97**	**2.46**	**0.66**
	0.200	0.20	0.20	0.19	0.20	0.500	0.50	0.50	0.49	0.50	0.800	0.80	0.79	0.76	0.80
								$T = 15$							
30	**0.049**	**2.51**	**0.99**	**1.61**	**1.01**	**0.051**	**2.45**	**1.01**	**2.31**	**0.96**	**0.054**	**1.90**	**1.01**	**4.02**	**0.78**
	0.200	0.15	0.19	0.15	0.20	0.490	0.49	0.49	0.42	0.50	0.800	0.79	0.78	0.65	0.80
300	**0.015**	**0.97**	**1.00**	**1.39**	**0.99**	**0.017**	**1.16**	**0.99**	**1.45**	**0.93**	**0.016**	**1.16**	**1.02**	**2.73**	**0.83**
	0.200	0.20	0.20	0.19	0.20	0.500	0.50	0.50	0.49	0.50	0.800	0.80	0.80	0.77	0.80

For the FOSLS columns we report 1.4826*MAD(FOSLS) to approximate its RMSE. The columns for other estimators give the relative MADs compared to MAD(FOSLS). The relative MAD is typed in bold and the Median is typed in a regular font.

Table 2.2: Relative MAD and Median of the Estimators (Nonnorrmal Stationary)

	$\alpha_0 = 0.2$					$\alpha_0 = 0.5$					$\alpha_0 = 0.8$				
	FOSLS	FOSLS1	RML	GMM	MPML	FOSLS	FOSLS1	RML	GMM	MPML	FOSLS	FOSLS1	RML	GMM	MPML
N								$T = 5$							
30	**0.109**	**1.06**	**1.06**	**1.47**	**1.13**	**0.120**	**1.02**	**1.11**	**1.93**	**1.18**	**0.133**	**1.11**	**1.44**	**4.23**	**1.07**
	0.190	0.18	0.18	0.14	0.21	0.490	0.46	0.45	0.36	0.49	0.780	0.72	0.68	0.42	0.79
300	**0.035**	**1.02**	**1.06**	**1.45**	**1.10**	**0.042**	**1.07**	**1.14**	**1.67**	**1.12**	**0.071**	**1.01**	**1.12**	**2.53**	**0.67**
	0.200	0.20	0.20	0.19	0.20	0.500	0.50	0.49	0.48	0.50	0.800	0.80	0.78	0.70	0.80
								$T = 10$							
30	**0.070**	**0.95**	**0.93**	**1.41**	**0.97**	**0.066**	**1.02**	**1.02**	**2.07**	**1.08**	**0.074**	**1.07**	**1.05**	**3.98**	**0.94**
	0.190	0.19	0.19	0.15	0.20	0.490	0.49	0.48	0.41	0.49	0.790	0.77	0.76	0.60	0.79
300	**0.022**	**0.89**	**1.00**	**1.25**	**0.98**	**0.022**	**1.08**	**1.03**	**1.48**	**1.14**	**0.026**	**1.04**	**1.07**	**2.50**	**0.98**
	0.200	0.20	0.20	0.19	0.20	0.500	0.50	0.50	0.49	0.50	0.800	0.80	0.79	0.76	0.80
								$T = 15$							
30	**0.049**	**1.40**	**1.00**	**1.53**	**0.98**	**0.049**	**1.60**	**1.11**	**2.21**	**1.08**	**0.048**	**1.50**	**1.07**	**4.35**	**1.11**
	0.200	0.18	0.19	0.15	0.20	0.490	0.49	0.49	0.43	0.50	0.790	0.78	0.78	0.66	0.79
300	**0.017**	**0.81**	**0.95**	**1.19**	**0.99**	**0.016**	**1.05**	**1.01**	**1.51**	**1.15**	**0.017**	**1.02**	**1.06**	**2.66**	**1.12**
	0.200	0.20	0.20	0.20	0.20	0.500	0.50	0.50	0.49	0.50	0.800	0.80	0.80	0.77	0.80

For the FOSLS columns we report 1.4826*MAD(FOSLS) to approximate its RMSE. The columns for other estimators give the relative MADs compared to MAD(FOSLS). The relative MAD is typed in bold and the Median is typed in a regular font.

Table 2.3: Relative MAD and Median of the Estimators (Normal Nonstationary)

	$\alpha_0 = 0.2$					$\alpha_0 = 0.5$					$\alpha_0 = 0.8$				
	FOSLS	FOSLS1	RML	GMM	MPML	FOSLS	FOSLS1	RML	GMM	MPML	FOSLS	FOSLS1	RML	GMM	MPML
N								$T = 5$							
30	**0.012**	**1.14**	**1.01**	**0.99**	**89.71**	**0.007**	**1.16**	**0.96**	**0.97**	**89.11**	**0.005**	**1.07**	**0.97**	**0.97**	**51.86**
	0.200	0.20	0.20	0.20	0.90	0.500	0.50	0.50	0.50	0.94	0.800	0.80	0.80	0.80	0.96
300	**0.004**	**1.04**	**1.01**	**1.00**	**283.35**	**0.002**	**1.04**	**1.00**	**1.00**	**269.12**	**0.001**	**1.01**	**0.99**	**0.99**	**172.63**
	0.200	0.20	0.20	0.20	0.91	0.500	0.50	0.50	0.50	0.94	0.800	0.80	0.80	0.80	0.96
								$T = 10$							
30	**0.011**	**1.35**	**1.03**	**1.02**	**103.03**	**0.006**	**1.49**	**0.98**	**0.96**	**119.27**	**0.002**	**1.52**	**0.98**	**0.98**	**121.27**
	0.200	0.20	0.20	0.20	0.95	0.500	0.50	0.50	0.50	0.97	0.800	0.80	0.80	0.80	0.99
300	**0.004**	**0.99**	**1.00**	**1.00**	**317.01**	**0.002**	**1.11**	**0.98**	**0.99**	**389.06**	**0.001**	**1.10**	**1.01**	**1.00**	**368.02**
	0.200	0.20	0.20	0.20	0.95	0.500	0.50	0.50	0.50	0.97	0.800	0.80	0.80	0.80	0.99
								$T = 15$							
30	**0.011**	**1.86**	**0.98**	**0.98**	**105.15**	**0.005**	**2.38**	**0.98**	**0.98**	**133.17**	**0.002**	**2.25**	**0.98**	**0.95**	**149.49**
	0.200	0.20	0.20	0.20	0.97	0.500	0.50	0.50	0.50	0.98	0.800	0.80	0.80	0.80	0.99
300	**0.003**	**1.01**	**1.00**	**1.02**	**330.53**	**0.002**	**1.09**	**0.98**	**0.99**	**409.54**	**0.001**	**1.17**	**1.00**	**1.01**	**512.12**
	0.200	0.20	0.20	0.20	0.97	0.500	0.50	0.50	0.50	0.98	0.800	0.80	0.80	0.80	0.99

For the FOSLS columns we report 1.4826∗MAD(FOSLS) to approximate its RMSE. The columns for other estimators give the relative MADs compared to MAD(FOSLS). The relative MAD is typed in bold and the Median is typed in a regular font.

Table 2.4: Relative MAD and Median of the Estimators (Nonnormal Nonstationary)

	$\alpha_0 = 0.2$					$\alpha_0 = 0.5$					$\alpha_0 = 0.8$				
	FOSLS	FOSLS1	RML	GMM	MPML	FOSLS	FOSLS1	RML	GMM	MPML	FOSLS	FOSLS1	RML	GMM	MPML
N								$T=5$							
30	**0.009**	**1.07**	**1.29**	**1.29**	**120.57**	**0.005**	**1.07**	**1.44**	**1.44**	**124.24**	**0.003**	**1.18**	**1.56**	**1.54**	**89.56**
	0.200	0.20	0.20	0.20	0.90	0.500	0.50	0.50	0.50	0.94	0.800	0.80	0.80	0.80	0.96
300	**0.003**	**0.93**	**1.44**	**1.44**	**394.89**	**0.002**	**0.95**	**1.48**	**1.47**	**413.21**	**0.001**	**0.89**	**1.44**	**1.44**	**255.54**
	0.200	0.20	0.20	0.20	0.91	0.500	0.50	0.50	0.50	0.94	0.800	0.80	0.80	0.80	0.96
								$T=10$							
30	**0.008**	**1.26**	**1.38**	**1.34**	**146.48**	**0.004**	**1.32**	**1.55**	**1.52**	**191.90**	**0.002**	**1.16**	**1.46**	**1.45**	**173.13**
	0.200	0.20	0.20	0.20	0.95	0.500	0.50	0.50	0.50	0.97	0.800	0.80	0.80	0.80	0.99
300	**0.002**	**0.95**	**1.44**	**1.47**	**468.00**	**0.001**	**0.95**	**1.47**	**1.46**	**554.61**	**0.000**	**1.01**	**1.53**	**1.53**	**590.03**
	0.200	0.20	0.20	0.20	0.95	0.500	0.50	0.50	0.50	0.97	0.800	0.80	0.80	0.80	0.99
								$T=15$							
30	**0.007**	**2.00**	**1.37**	**1.40**	**156.16**	**0.004**	**2.38**	**1.43**	**1.42**	**193.59**	**0.001**	**2.18**	**1.53**	**1.54**	**239.64**
	0.200	0.20	0.20	0.20	0.97	0.500	0.50	0.50	0.50	0.98	0.800	0.80	0.80	0.80	0.99
300	**0.002**	**1.06**	**1.50**	**1.52**	**497.46**	**0.001**	**1.10**	**1.56**	**1.57**	**641.36**	**0.000**	**1.18**	**1.59**	**1.60**	**748.73**
	0.200	0.20	0.20	0.20	0.97	0.500	0.50	0.50	0.50	0.98	0.800	0.80	0.80	0.80	0.99

For the FOSLS columns we report 1.4826∗MAD(FOSLS) to approximate its RMSE. The columns for other estimators give the relative MADs compared to MAD(FOSLS). The relative MAD is typed in bold and the Median is typed in a regular font.

rate than the FOSLS, which was expected. The downward bias in FOSLS vanishes quickly as T increases.

Table 2.2 shows the wide out-performance of the FOSLS for small and large N and specially for small T. The relative MADs of RML reveal that the true levels of variance reduction gained by FOSLS (see Figure 2.1.a) require N to be larger than 300. The FOSLS competes well with the MPML for small N. Although the FOSLS1 is not reliable for small ratio N/T, it performs well for $N = 300$. The FOSLS has smaller bias for small N than other estimators.

Table 2.3 shows that the FOSLS, RML and GMM compete very well for small and large N. The close performance of the FOSLS and RML is due to the normality of ε_{it}. The improvement in the GMM performances is due to the nonstationary of the y_{it} process. The table shows how the MPML breaks down everywhere under this scenario. The FOSLS1 requires large ratio N/T to get stable so it is not recommended in this scenario.

Table 2.4 shows the effect of the skewness of the ε_{it} distribution on the performance of the FOSLS. The RML and GMM are less efficient than the FSOLS by at least 30% for $N = 30$, and by as high as 59% for $N = 300$. The relative MADs of the RML for large N are consistent with Figure (2.1.c). Although the true levels of variance reduction gained by FOSLS require N to be larger than 300, yet the gain of efficiency for small N is much larger than the corresponding gain in the nonnormal stationary scenario.

The numerical results of the three remaining nonnormal nonstationary scenarios (b,c,d) show that the outperformance of the FOSLS(α_0) appears clearly if ε_{it} is not normally distributed. .

Performance in the Presence of Covariates

In the previous simulation studies we considered models without covariates in order to make our numerical results comparable with existing results in the literature such as (Blundell and Bond, 1998; Alvarez and Arellano, 2003; Okui, 2009). In what follows we investigate if the finite sample performance of the FOSLS will change if we add some covariates in the model. In particular, we consider the data generating process of (Kiviet, 1995, Appendix B), who compared some least squares and instrumental variable (IV) type estimators (among them the linear first difference GMM of Arelano and Bond 1991) over more than 14 parameters combinations (designs). Following (Kiviet, 1995, Section 5 and Appendix B) and using his notations and definitions, the structural equation for the individual i takes the form (we omit the subscript i)

$$y_t = \gamma y_{t-1} + \beta x_t + \eta + \varepsilon_t, \quad t = 1, 2, \ldots, T, \tag{2.11}$$

and the y_t-process is generated from the reduced equation (Kiviet, 1995, B8)

$$y_t = \beta\phi_t + \psi_t + \eta/(1-\gamma), \quad t = 0, 1, 2, \ldots, T, \tag{2.12}$$

where $\phi_t \sim AR(2)$ and $\psi_t \sim AR(1)$ are mutually independent stationary processes and both are independent of $\eta \sim N(0, \sigma_\eta^2)$.

The orthogonality and normality assumptions of Kiviet (1995) imply the following identities:

$$E(\eta\, x_t) = 0 \quad t = 0, 1, 2, \ldots, T, \tag{2.13}$$

$$V(y_0) = \beta^2 V(\phi_0) + V(\psi_0) + \sigma_\eta^2/(1-\gamma)^2, \tag{2.14}$$

$$E(\eta\, y_0) = \sigma_\eta^2/(1-\gamma), \tag{2.15}$$

$$E(x_t\, y_0) = \lambda\, \rho^t \quad t = 0, 1, 2, \ldots, T, \quad \lambda \equiv \beta\, \sqrt{V(\phi_0)V(\xi_0)/(1-\rho^2)} \tag{2.16}$$

and

$$E(\eta | y_0, x_1, \ldots, x_T) = f_1(y_0, x_1, \ldots, x_T, \boldsymbol{\theta}) = \theta_2\, y_0 + \theta_4\, x_1, \tag{2.17}$$

where

$$\theta_2 = \frac{\sigma_\eta^2}{(1-\gamma)(V(y_0) - \lambda^2(1-\rho^2)/V(\xi_0))} \tag{2.18}$$

and

$$\theta_4 = \theta_2\, \lambda\, \rho\, (1-\rho^2)/V(\xi_0). \tag{2.19}$$

Moreover, the joint normality implies that $V(\eta | y_0, x_1, \ldots, x_T)$ is constant and does not depend on $y_0, x_1, \ldots, x_T$, and hence the conditional variance is represented in terms of one extra parameter, namely $\exp(\theta_3)$ as in our notation. Accordingly we have totally seven parameters. Tables 2.5 and 2.6 give summary statistics (bias, standard deviation, root mean squred error) of the linear first differenced GMM1 (Arellano and Bond, 1991), RML, and FOSLS across the 14 mentioned designs.

As expected, the RML (which is the conditional MLE) is the best in all cases because the data are generated using the normal distribution. However, the FOSLS (which is asymptotically as efficient as the RML) is very close to the RML, as can be seen from Tables 2.5 and 2.6. These results show that in the presence of strictly

Table 2.5: Bias, Standard deviation and RMSE of FOSLSE for dynamic linear model (compared to Kiviet 1995), $N = 100$, $T = 6$

		Bias		Std		RMSE	
Design	Estimator	α	β	α	β	α	β
I	GMM1	-0.013	0.003	0.166	0.120	0.168	0.120
	RML	0.000	0.013	0.092	0.087	0.092	0.087
	FOSLS	0.000	0.013	0.110	0.104	0.112	0.104
II	GMM1	-0.029	0.002	0.388	0.130	0.396	0.131
	RML	-0.005	-0.006	0.106	0.093	0.107	0.094
	FOSLS	-0.005	-0.006	0.107	0.098	0.107	0.099
III	GMM1	-0.060	-0.002	0.215	0.610	0.216	0.611
	RML	-0.020	0.008	0.117	0.405	0.117	0.405
	FOSLS	-0.019	0.007	0.141	0.509	0.146	0.509
IV	GMM1	-0.015	0.007	0.358	0.616	0.361	0.616
	RML	-0.003	0.019	0.146	0.436	0.146	0.436
	FOSLS	-0.003	0.019	0.153	0.503	0.154	0.503
V	GMM1	-0.038	-0.007	0.126	0.057	0.128	0.057
	RML	-0.003	-0.015	0.057	0.042	0.058	0.043
	FOSLS	-0.003	-0.014	0.067	0.047	0.068	0.047
VI	GMM1	-0.060	-0.026	0.304	0.065	0.313	0.066
	RML	-0.021	0.023	0.059	0.046	0.059	0.047
	FOSLS	-0.020	0.023	0.063	0.047	0.063	0.048

GMM1 stands for the linear first differenced GMM as in (Arellano and Bond, 1991).

There is little difference between our GMM1 results and Kiviet's results. This can be explained in light of two mistakes we found in kiviet's equation B6, namely ξ_0 and ξ_1 are supposed to be standardized before used. Some correspondences were done with Kiviet about this issue.

Table 2.6: Bias, Standard deviation and RMSE of FOSLSE for dynamic linear model (compared to Kiviet 1995), $N = 100$, $T = 3$

		Bias		Std		RMSE	
Design	Estimator	α	β	α	β	α	β
VII	GMM1	-0.024	-0.008	0.277	0.291	0.280	0.292
	RML	-0.001	0.009	0.112	0.195	0.112	0.195
	FOSLS	-0.021	0.010	0.134	0.246	0.138	0.246
VIII	GMM1	-0.026	-0.005	0.388	0.297	0.392	0.298
	RML	-0.006	0.004	0.139	0.210	0.140	0.211
	FOSLS	-0.010	0.005	0.147	0.236	0.147	0.236
IX	GMM1	-0.081	-0.014	0.056	0.071	0.057	0.071
	RML	0.006	0.017	0.036	0.050	0.036	0.052
	FOSLS	-0.002	0.015	0.037	0.050	0.037	0.052
X	GMM1	-0.074	-0.011	0.065	0.309	0.067	0.309
	RML	-0.004	0.009	0.048	0.182	0.048	0.183
	FSOLS	-0.005	0.009	0.048	0.183	0.048	0.184
XI	GMM1	-0.027	-0.036	0.072	0.069	0.077	0.069
	RML	-0.001	0.006	0.044	0.045	0.044	0.045
	FOSLS	-0.036	0.008	0.044	0.045	0.044	0.046
XII	GMM1	-0.038	-0.019	0.088	0.344	0.096	0.344
	RML	0.000	0.006	0.058	0.178	0.058	0.179
	FOSLS	-0.030	0.003	0.058	0.180	0.058	0.180
XIII	GMM1	-0.047	-0.034	0.099	0.148	0.116	0.148
	RML	0.013	0.012	0.056	0.086	0.059	0.086
	FOSLS	-0.006	-0.014	0.059	0.087	0.062	0.087
XIV	GMM1	-0.051	-0.018	0.100	0.877	0.117	0.877
	RML	0.013	0.009	0.057	0.430	0.060	0.431
	FOSLS	-0.005	0.007	0.060	0.434	0.063	0.435

GMM1 stands for the linear first difference GMM as in (Arellano and Bond, 1991).
There is little difference between our GMM1 results and Kiviet's results. This can be explained in light of two mistakes we found in kiviet's equation B6, namely ξ_0 and ξ_1 are supposed to be standardized before used. Some correspondences were done with Kiviet about this issue.

exogenous variable the FOSLS is a good alternative to the most efficient RML under normality.

We have also done another simulation under nonnormal nonstationary setup (including strictly exogenous variable) which is different from the setup of Kiviet (1995). The results are not reported because they are similar to those in Table 2.4. They also emphasize the outperformance of FOSLS compared to the RML due to deviation from normality.

Robustness of FOSLS against Near Unit-root Specification

It is well known in the literature that linear first-differenced GMM is generally inferior when the true value of the autoregressive parameter; α_0 is close to one due to the problem of weak instruments see (Blundell and Bond, 1998). On the other hand, from Tables 2.1–2.4 we have seen that the performance of FOSLS is not affected by large values of α_0. Furthermore, it becomes closer to the MPML and sometimes is even better due to the efficiency gain in the case of nonnormal disturbances, see Table 2.2. Hence it is interesting to investigate the performance of FOSLS when α_0 is very close to one (unit root case) and see if it breaks down like the GMM or it is stable like the MPML.

To answer these questions we carry out a simulation using the nonnormal stationary process defined on page 53. The sample sizes are $N = 100$, 300 and $T =$ 3, 6 and parameter values are $\alpha_0 = 0.9$, 0.95, 0.99. The results are reported in Table 2.7 which shows that the FOSLS is more efficient than the RML due to the skeweness in the within group disturbance and it comes with less bias. Although the FSOLS has larger downward bias than the MPML but the difference vanishes

Table 2.7: Relative RMSE and Mean of the Estimators (Nonnorrmal Near Unit Root)

	$\alpha_0 = 0.9$				$\alpha_0 = 0.95$				$\alpha_0 = 0.99$			
	FOSLS	RML	GMM	MPML	FOSLS	RML	GMM	MPML	FOSLS	RML	GMM	MPML
N						$T = 3$						
100	**0.216**	**1.23**	**4.94**	**0.45**	**0.240**	**1.13**	**5.04**	**0.38**	**0.123**	**2.36**	**10.28**	**0.66**
	0.82	0.71	0.30	0.88	0.87	0.75	0.12	0.92	0.95	0.77	0.08	0.95
300	**0.151**	**1.21**	**5.85**	**0.43**	**0.130**	**1.52**	**8.32**	**0.44**	**0.060**	**3.59**	**21.85**	**0.81**
	0.83	0.78	0.48	0.90	0.89	0.81	0.27	0.94	0.97	0.83	0.09	0.97
						$T = 6$						
100	**0.086**	**1.27**	**5.27**	**0.67**	**0.094**	**1.32**	**6.21**	**0.49**	**0.037**	**3.75**	**17.10**	**1.07**
	0.87	0.83	0.52	0.90	0.91	0.86	0.43	0.94	0.98	0.88	0.43	0.97
300	**0.039**	**1.85**	**7.44**	**0.88**	**0.064**	**1.36**	**7.83**	**0.48**	**0.013**	**7.50**	**47.90**	**1.69**
	0.90	0.86	0.68	0.90	0.92	0.89	0.53	0.95	0.99	0.91	0.43	0.98

For the FOSLS columns we report its RMSE. The columns for other estimators give the relative RMSEs compared to RMSE(FOSLS). The relative MSE is typed in bold and the Mean is typed in a regular font.

quickly as T and α_0 increase. These results demonstrate that the FSOLS is robust against unit root process and is even more efficient than the MPML for $T = 6$ and $\alpha_0 = 0.99$. This is the case because, first, the MPML loses some efficiency due to nonnormality and, second, the MPML is not allowed to go over the boundary 1. In contrast, the FSOLS is computed in a fashion which permits so. Generally speaking, the FSOLS gets closer to the MPML as α_0 increases.

We conclude this section by examining to what extent the asymptotic variance formula for the OSLS (see A.12) can be used to approximate its counterpart for different values of N, T and α_0. Table 2.8 shows that the asymptotic formula provides reliable estimates of the corresponding small sample standard errors (as given by the simulations). In the next section we use the U.S. airfares data to apply the FOSLS and investigate its practical merits.

Table 2.8: The RMSE of the FOSLS calculated by simulation and using the asymptotic formula (Nonnormal Nonstationary)

	$\alpha_0 = 0.2$		$\alpha_0 = 0.5$		$\alpha_0 = 0.8$	
	Simulation	Asy Formula	Simulation	Asy Formula	Simulation	Asy Formula
N			$T = 5$			
30	0.00866	0.00815	0.00526	0.00501	0.00268	0.00288
300	0.00265	0.00258	0.00159	0.00158	0.00094	0.00091
			$T = 10$			
30	0.00761	0.00724	0.00365	0.00388	0.00161	0.00157
300	0.00239	0.00229	0.00127	0.00123	0.00047	0.00050
			$T = 15$			
30	0.00729	0.00703	0.00369	0.00364	0.00119	0.00122
300	0.00229	0.00222	0.00112	0.00115	0.00038	0.00039

2.5 Empirical Example: The U.S. Airfare

In this section we use a real data example to demonstrate the practical usefulness of the SLS approach in comparison with the IV approach and to assess the practical gain of efficiency over the RML estimator. In particular, we use a data set published in (Wooldridge, 2010) and downloadable from `http://mitpress.mit.edu/sites/default/files/titles/content/wooldridge/statafiles.zip`. The dataset airfare.dta contains data for the average airfares, average number of passengers, distance, and the market share of the largest carrier for each of the top 1149 city-pair markets within the contiguous 48 US states for the fourth quarters of 1997, 1998, 1999, and 2000. The data are from the Domestic Airline Fares Consumer Report published by the U.S. Department of Transportation. The city pairs in the sample account for about 75 percent of total within-48-state passenger trips. The panels are balanced. A fairly clear description of the data can be found at website `http://academic.reed.edu/economics/parker/s10/312/Asgns/proj3.html`.

Among the main factors that are likely to influence airfares are the flight distance (directly related to flight-crew wages, fuel and airplane user cost), the market size (average number of passengers per day flying the route), and market concentration (fraction of route traffic accounted for by the largest carrier). The ticket prices are set and adjusted through a fairly dynamic and interactive system. In addition, most airfares determinants are also influenced by the airfare hence they are endogenous variables. Since in most cases the airline set the current airfare by adjusting the previous year fare, a linear dynamic panel regression model with unobserved route heterogeneity and time dummies may be appropriate to measure the dynamic effect of those determinants. All the variables (except the concentration) are measured in logarithmic scale and hence we are estimating the elasticity of the airfare with respect to other determinants.

The bivariate scatterplots are shown in Figure 2.2. It shows linear dependence

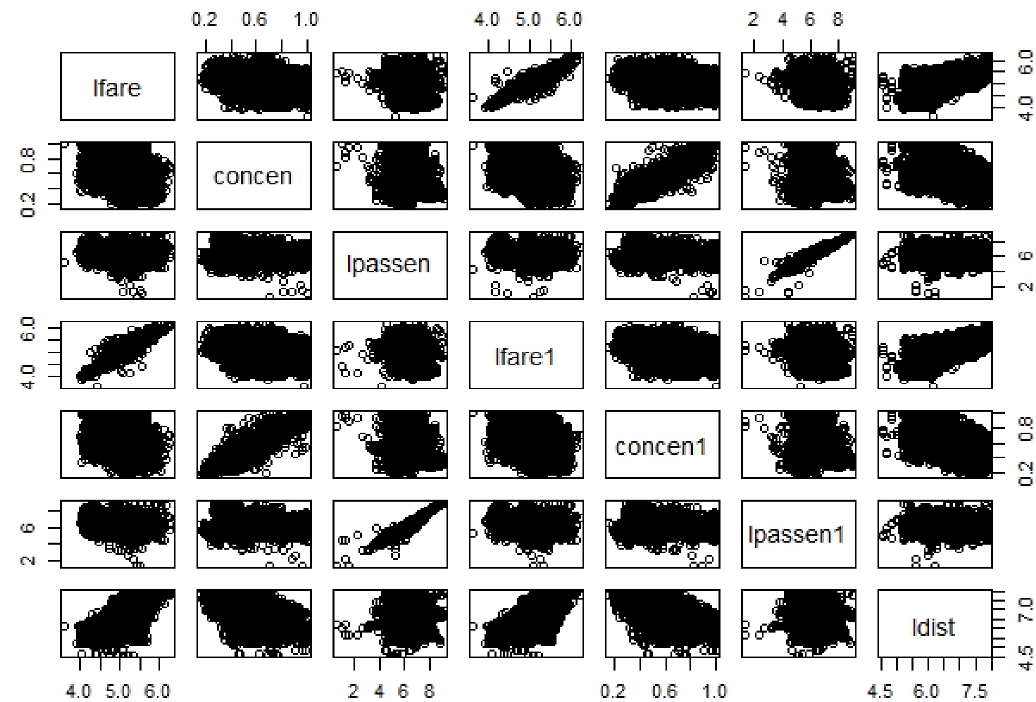


Figure 2.2: Matrix plot for the airfares determinants

between the current and previous (log of) airfares; lfare and lfare1 respectively, justifying a dynamic linear panel data model. Matching with rational expectation Figure 2.2 shows moderate positive linear association between the lfare and ldist. Another factor that is likely to influence airfares is the amount of competition on the route. More concentrated routes would be expected to have higher fares than routes on which many airlines compete. A higher value of the variable concen should be associated with higher lfare. However, this is not clear from Figure 2.2. A possible explanation is the clear negative correlation between the concen and ldist which -given the positive correlation between the lfare and ldist- probably hides the rational positive correlation between concen and lfare. The relation between lfare and the first lag of concen (concen1) is not clear either theoretical or empirically. It can be positive or negative depending on how fast the airlines market can change from the state of monopoly to the state of perfect competition over the time.

Full planes are cheaper (per passenger) than empty ones, so costs, and therefore fares, are expected to be lower on heavily travelled routes. More serious, the log of average number of passengers (lpassen) is likely to be endogenous to the determination of airfares. Figure 2.2 shows weak correlation between (lpassen or lpassen1) and the lfare. This again emphasizes the need for multiple regression model. Although we are expecting to see negative correlation between the lpassen and lfare, the direction of the relationship between lpassen1 and lfare can be positive or negative depending on how the airlines set their ticket prices in response to last year changes in average number of passengers per day flying the route.

In light of the above discussion, it reasonable to start by a linear dynamic model with unobserved route heterogeneity and time dummies $D99$, $D00$ for years 1999 and 2000 respectively. The model equation takes the form

$$\begin{aligned} lfare_{it} = \alpha_0 + \alpha_1\, lfare_{i(t-1)} + \beta_1\, ldist_i + \beta_2\, concen_{it} + \beta_3\, concen_{i(t-1)} + \\ \beta_4\, lpassen_{it} + \beta_5\, lpassen_{i(t-1)} + \beta_5 D99 + \beta_6\, D00 + \eta_i + \varepsilon_{it} \\ i = 1, 2, \ldots, 1149, \quad t = 1998, 1999, 2000. \quad (2.20) \end{aligned}$$

We started by fitting the proposed equation (2.20) using a naive OLS estimation. The results are shown in the OLS column of Table 2.9 (page 71). It is well known that the OLSEs are not consistent due to the built-in correlation between η_i and $lfare_{i(t-1)}$, however the high goodness of fit of the OLS fitted model ($R^2 = 0.95$) reflects the high explanatory power in the regressors. We also use the two stage linear first differenced (GMM2) of Arellano and Bond (1991) to fit the model, where the lags of order at least two of $lfare$, $concen$, and $lpassen$ are used as instrumental variables. This was necessary to turn around the possible endogeneity of (concen, lpassen) and of course lfare. We used the xtdpd command in STATA 13.0:

```
xtdpd L(0/1).lfare y99 y00 L(0/1).concen L(0/1).lpassen,
noconstant twostep dgmmiv(lfare) dgmmiv(concen) dgmmiv(lpassen)
div(y99) div(y00) div(ldist, nodifference) artests(2)
```

The results are shown in Table 2.9 under Model-I. Wooldridge (2010, 373) used the GMM2 to fit a first order linear dynamic model which includes only concen (treated as strictly exogenous) in addition to the dummies. The estimate of the autoregressive parameter was 0.333 which is not far from ours. According to the reported value of Sargan test (see Table 2.9) the GMM sequential moments based on the reduced difference form of model (2.20) are not rejected. The only concern is the negative sign of the estimated elasticity of lfare with respect to concen (p-value=0.059). A possible explanation for this unexpected sign is the problem of weak instruments which is likely to occur in identifying β_2 given the strong positive linear correlation between concen and concen1 as seen in Figure 2.2. This is similar to the situation where we use the first differenced GMM estimator to estimate the

Table 2.9: The models fitted to the U.S. Airfare data from 1997 through 2000.

	Model-I				Model-II			Model-III			Model-IV		
Coef.	OLS	GMM2	RML	FOSLS	GMM2	RML	FOSLS	GMM2	RML	FOSLS	GMM2	RML	FOSLS
Const: θ_0	**0.153**	—	**0.256**	**-0.122**	—	**0.248**	**0.723**	—	**0.243**	**0.786**	—	**0.279**	**0.708**
	—	—	0.032	0.032	—	0.026	0.025	—	0.010	0.010	—	0.051	0.050
lfare1: α_1	**0.929**	**0.216**	**0.374**	**0.536**	**0.216**	**0.377**	**0.461**	**0.157**	**0.373**	**0.817**	**0.169**	**0.147**	**0.823**
	—	0.100	0.039	0.033	0.100	0.014	0.013	0.080	0.009	0.009	0.072	0.017	0.017
ldist: β_1	**0.031**	—	**0.049**	**0.066**	—	**0.051**	**0.091**	—	**0.052**	**-0.007**	—	**0.062**	**-0.019**
	—	—	0.004	0.004	—	0.003	0.003	—	0.001	0.001	—	0.006	0.006
concen: β_2	**0.119**	**-0.849**	**0.108**	**0.048**	**-0.849**	**0.091**	**0.162**	**-0.217**	**0.080**	**0.070**	**-0.842**	**0.095**	**0.077**
	—	0.449	0.018	0.018	0.449	0.011	0.011	0.167	0.003	0.003	0.389	0.020	0.019
concen1 β_3	**-0.076**	**0.325**	**0.013**	**0.040**	**0.325**	**-0.015**	**-0.032**	—	—	—	**0.320**	**0.000**	**-0.106**
	—	0.215	0.022	0.022	0.215	0.011	0.011	—	—	—	.187	0.020	0.019
lpassen: β_4	**-0.374**	**-0.404**	**-0.367**	**-0.556**	**-0.404**	**-0.367**	**-0.108**	**-0.450**	**-0.367**	**-0.094**	**-0.382**	**-0.333**	**-0.084**
	—	0.081	0.006	0.006	0.081	0.004	0.004	0.057	0.002	0.002	0.060	0.007	0.007
lpassen1: β_5	**0.374**	**0.103**	**0.175**	**0.427**	**0.103**	**0.176**	**-0.014**	**0.106**	**0.175**	**0.005**	—	—	—
	—	0.141	0.021	0.019	0.141	0.006	0.006	0.111	0.003	0.003	—	—	—
D99: β_6	**0.000**	**0.003**	**0.016**	**0.009**	**0.003**	**0.016**	**0.037**	**0.020**	**0.016**	**-0.017**	**0.005**	**0.023**	**-0.014**
	—	0.013	0.003	0.003	0.013	0.002	0.002	0.006	0.001	0.001	0.012	0.003	0.003
D00: β_7	**0.042**	**0.070**	**0.074**	**0.060**	**0.070**	**0.073**	**0.092**	**0.087**	**0.073**	**-0.003**	**0.078**	**0.090**	**0.073**
	—	0.015	0.003	0.003	0.015	0.0026	0.002	0.011	0.001	0.001	0.012	0.003	0.003
						Diagnostic Statistics							
SW/Sargan		**6.41**		**20.98**	**6.41**		**4.01**	**13.90**		**335.80**	**8.16**		**8545.16**
P-value		0.268		0.013	0.268		0.911	0.031		0.000	0.226		0.000
RMSE	0.09	5.12	4.02	3.23	5.12	4.04	2.66	5.12	4.07	0.69	5.12	6.01	0.74

The coefficients estimates are typed in bold and their standard errors are typed in regular font. The standard errors of RML and FOSLS are computed using formula (A.11 and A.12) respectively.

autoregressive parameter when it is closed to one (see Blundell and Bond, 1998). This problem causes inflation in the coefficient variance and may lead to unreliable estimates. This is an example where it is highly recommended to use the RE approach with the level data instead of adopting the FE approach with differenced data. The RE approach doesn't depend on instrumental variable and hence is able to give more reliable estimates in such a case.

However, since the sequential moments are correctly specified for Model-I, the GMM estimates can still be used to recover the within group errors ε_{it} and the unobserved routes effects η_i. We followed Arellano (2003, 118–119) to estimate the time effect for year 1998 and consequently η_i. Then realizations of ε_{it} are obtained directly from equation (2.20).

Examining the relationship between $\hat{\eta}_i$ and the initial values (1997) of the model variables, (lfare0, concen0, lpassen0 and ldsit0) will help in breaking the built-in correlation between the unobserved route effect and other regressors when we use the RML or the SLS approaches. Figure 2.3 (page 73) shows that η_i is likely to be linearly correlated with initial values. For more investigation on this point we fitted the following auxiliary regression equation

$$\hat{\eta}_i = -5.507 + 0.641\, lfare_{i0} + 0.464\, concen_{i0} - 0.041\, ldist_{i0} + 0.286\, lpassen_{i0}, \quad (2.21)$$

with all of its coefficients significant at 0.01 level of significance. Furthermore, Figure 2.4 (page 74) indicates that the variance of ε_{it} is time invariant. In addition, the within groups residuals don't show significant correlation over time. Hence the standard assumptions on η_i and ε_{it} can be adopted as reasonable working assumption for model (2.20). These assumptions will be tested implicitly when we use our SW

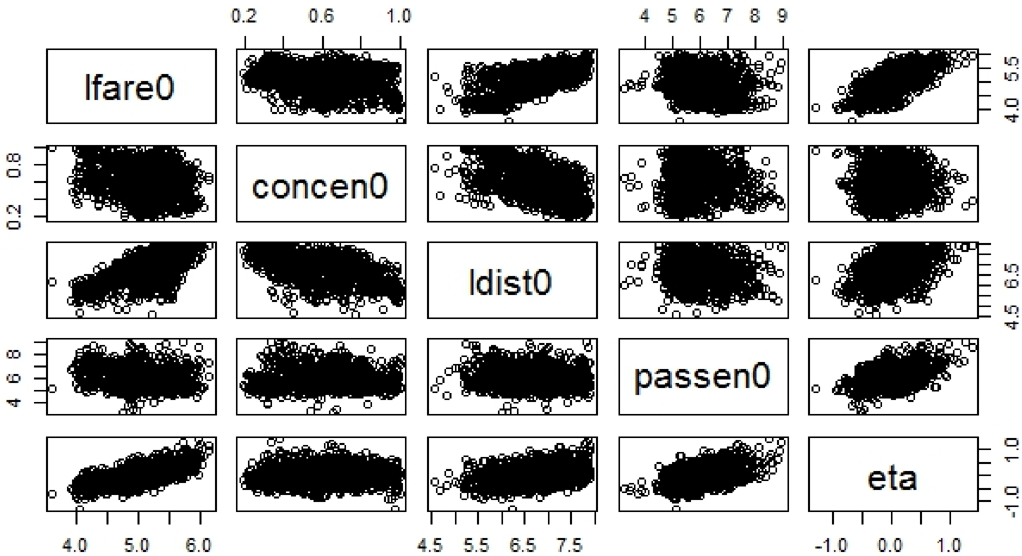


Figure 2.3: Correlation between GMM-$\hat{\eta}$ and initial values of the study variables.

test latter.

Based on the auxiliary equation (2.21) the RML estimates are obtained by fitting the following equation using the 'lme' command in R.

$$
\begin{aligned}
lfare_{it} = \theta_0 + \alpha_1\, lfare_{i(t-1)} + \beta_1\, ldist_i + \beta_2\, concen_{it} + \beta_3\, concen_{i(t-1)} + \\
\beta_4\, lpassen_{it} + \beta_5\, lpassen_{i(t-1)} + \beta_6 D99 + \beta_7\, D00 + \theta_1\, lfare_{i0} + \\
\theta_2\, concen_{i0} + \theta_3\, lpassen_{i0} + \eta_i^* + \varepsilon_{it}, \\
i = 1, 2, \ldots, 1149, \quad t = 1998, 1999, 2000. \quad (2.22)
\end{aligned}
$$

The coefficients estimates are given Table 2.9 under model-I. Figure 2.5 (page 75) shows the flat tail symmetrical distribution of the within group residuals $\hat{\varepsilon}_{it}$, which can be approximated by a Student t-distribution with df=5. Hence there is a possible

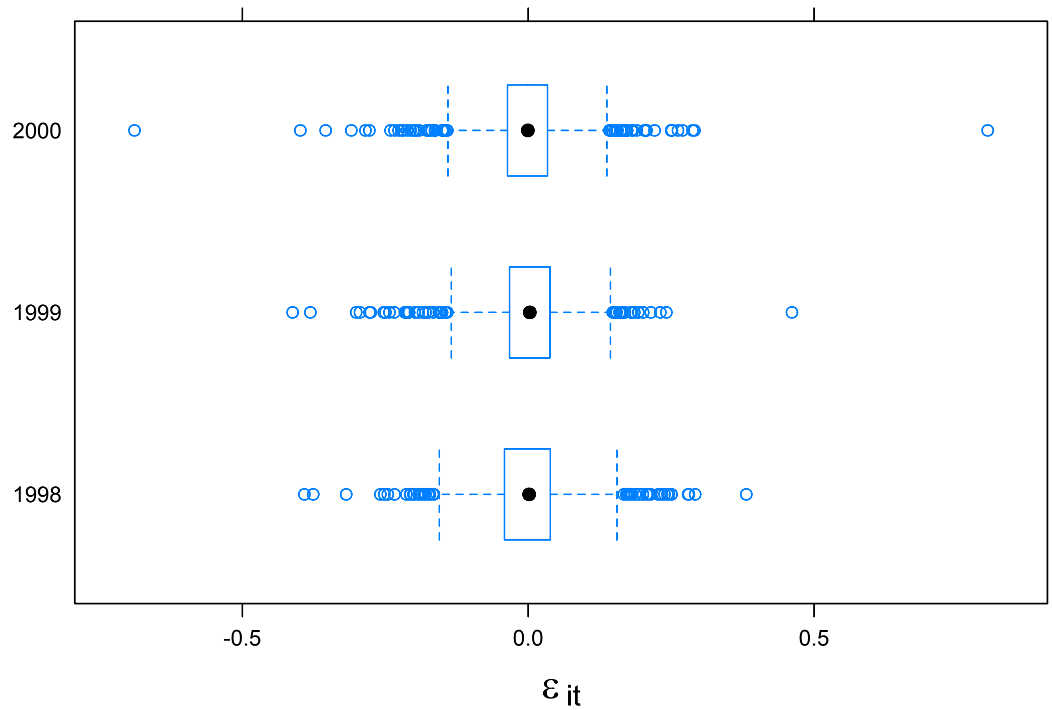


Figure 2.4: Homoscedasticity of GMM-$\hat{\varepsilon}_{it}$ across time

efficiency gain by applying the FOSLS estimator. We used the sample skewness and kurtosis of $\hat{\varepsilon}_{it}$, and $\hat{\eta}^*_{it}$ to calculate the optimal weight matrix and fit the FOSLS.

Only the estimates of the main regression equation are reported in Table 2.9. The estimates for the the other auxiliary variables lfare0, concen0, lpassen0 are not reported to save space and also the estimates of the variance components are not reported because they are not of main interest. The standard errors of the GMM2 are computed using the robust formula. The standard errors of RML and FOSLS reported in Table 2.9 are computed using formula (A.11 and A.12) respectively.

In Model-I, Sargan test doesn't provide evidence against the GMM sequential

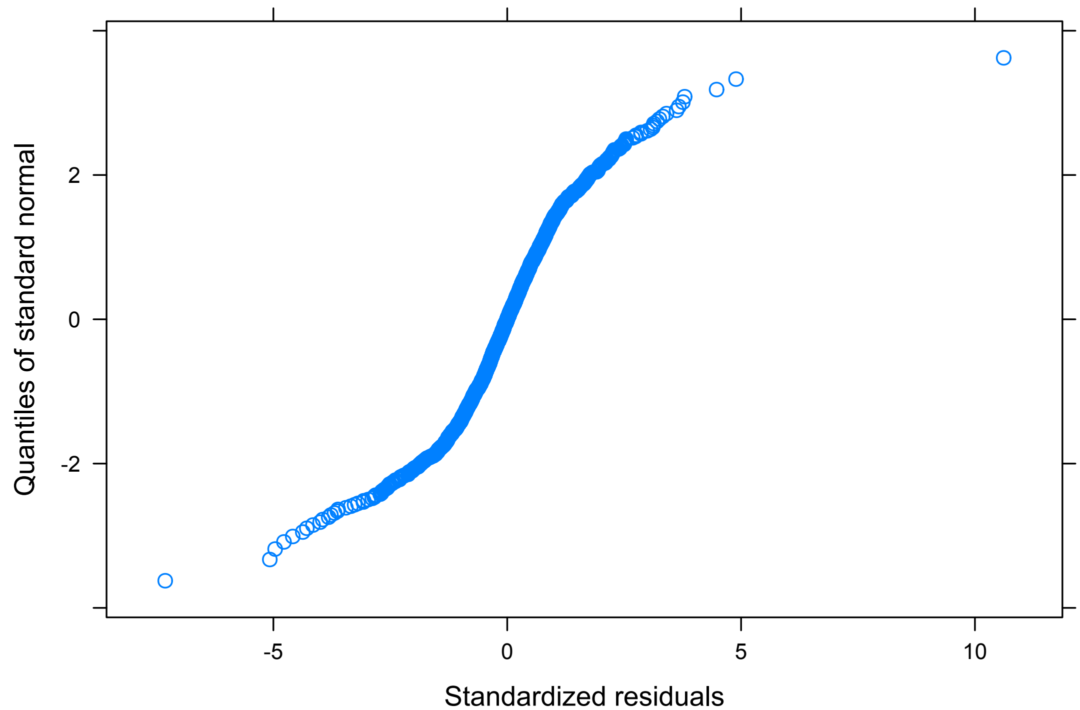


Figure 2.5: Flat tails symmetric distribution of RML-$\hat{\varepsilon}_{it}$

moments, however our test (SW) shows slight evidence against the first and second moments specification as given by equations (2.3 and 2.4). This motivates us to double check the empirical sampling distribution of SW under the specification defined by equations (2.3 and 2.4). We used the following steps: First, use the RML estimates of Model-I to generate data from equation (2.22), drawing η_i^* and ε_{it} from Student t-distribution with df 6 and 5 respectively to be as close as possible to the observed residuals and estimated random effect of Model-I. Second, use the generated data to fit equation (2.22) using the RML followed by FSOLS and then calculate the SW statistic. We Repeated these two steps 1000 times and plot a graph for the SW statistics to get approximation to the empirical sampling distribution of SW under H_0. Figure 2.6 shows that a sample size of 1149 routes is enough to

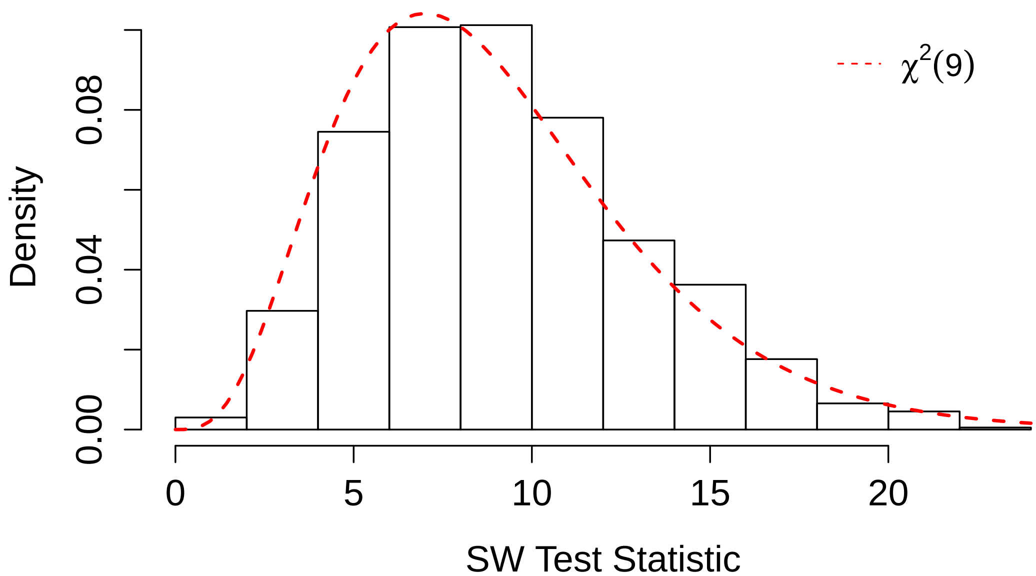


Figure 2.6: The empirical vs. asymptotic sampling distribution of SW using 1000 data cloning

make the empirical sampling distribution very close to the asymptotic one; χ^2_9 under H_0. Accordingly Model-I needs some adjustments by adding or eliminating some variables to pass the SW test.

To proceed, we fit equation (2.22) without the auxiliary variable concen0 because it is not of main interest and at the same time it has weak correlation with $\hat{\eta}_i$ as seen from Figure 2.3. This leads to Model-II which has the same list of explanatory variables as Model-I and of course the same value of Sargan test (6.41). On the other hand the value of SW statistic drops to 4.01 which is insignificant at any reasonable level of significance. This suggests that Model-II should be used for testing purpose on condition that the signs of FOSLS estimates are consistent with the economic theory. Before going further, it is important to take into consideration the robust standard error of the FSOLS as computed using formula in proposition 4. Due to lack of space in Table 2.9 they are reported in Table 2.10 below. According to these

Table 2.10: Model-II robust standard errors of the FOSLS calculated by proposition 4

θ_0 Const.	α_1 lfare1	β_1 ldist	β_2 concen	β_3 concen1	β_4 lpassen	β_5 lpassen1	β_6 D99	β_7 D00
0.061	0.041	0.007	0.034	0.034	0.026	0.023	0.004	0.005

robust standard errors, two variables are candidate to be dropped from the model, namely the concen1 and lpassen1.

Model-III is obtained by dropping only concen1 from model-II. However, both Sargan test and SW test reject this new specification. Accordingly, Model-IV is obtained by dropping only lpassen1 from model-II. Interestingly, SW test rejects Model-IV specification, while Sargan test doesn't reject the GMM sequential moments of this model. Since the GMM estimate of β_2 comes with negative sign (which doesn't make sense), model-IV is not recommended and therefore model-II is preferred. It does include both concen1 and lpassen1 although they may be insignificant if we use the robust standard errors to test for them. On the other hand, if we use the standard errors of RML and FOSLS as reported in Table 2.9 we can keep both concen1 and lpassen1 according to FOSLS and keep only lpassen1 according to RML. Since we know that Model-III is rejected by SW test, it follows that using the information inherent in the fourth moment through FOSLS was effective in keeping concen1 in model-II.

Finally, we emphasize that using the goodness of fit criteria for model selection may lead to completely inappropriate conclusion. For example, if we try to use the adjust R^2 or RMSE, we will select the OLS estimates which are completely misleading if we are interested in measuring the individual effects of the explanatory variables.

In summary, the outcomes of this empirical study are useful in emphasizing the following points.

First, differencing the data and using IVs is generally a risky approach to estimate linear dynamic panel models with covariates. Sometimes it produces misleading relations if the time varying variables are generated from autoregressive process with high autocorrelation (common situation with economic data). In such a case the linear first differenced GMM or other similar methods are not only unable to estimate the marginal effect of the time invariant variables but also, more harmful, weakly identify the marginal effect of the time variant explanatory variables.

Second, similar to Engel's empirical example in the first chapter the information inherent in the sample skewness and kurtosis of the within groups residuals can be used to gain some efficiency and consequently save some important variables from being wrongly eliminated from the model which may cause model misspecification. In other word, by using the extra efficiency of the OSLS we avoid falling in any misspecification traps.

Third, in this application we practically applied our proposed specification test, and examined its sampling distribution under the correct specification of the model using a data cloning approach. The empirical outcomes on this test cope with its theoretical ones, and we showed the usefulness of this test in selecting the suitable model instead of using other commonly used goodness of fit criteria (such as RMSE or R^2, AIC) which are completely misleading if the model is not correctly specified, especially with dynamic regression where the problem of endogenous is the default.

2.6 Summary

We examined SLS approach as an alternative to the commonly used RML or GMM estimators for the linear dynamic panel data model. The estimator is introduced through a semiparametric random effect approach which does not postulate any distributional assumptions on the error components in the model. This SLS is based on the first two moments of the outcome process given the initial observations and covariates. For large N and fixed T, the consistency and asymptotic normality of the SLS estimator were proven under fairly general regularity conditions. Furthermore, we showed that the optimal SLS efficiently uses the information given by the conditional moments of the outcome process. It reaches the bound of efficiency given by Chamberlain (1987).

Focusing on the autoregressive parameter α_0, we studied the over efficiency of the optimal SLS relative to the RML. We found that there is a considerable gain of efficiency when the distribution of the disturbances is asymmetric. From a computational point of view, we suggested two feasible versions of the optimal SLS. They (FOSLS, FOSLS1) are two-stage estimators having the same asymptotic properties as the true optimal SLS. Furthermore, we introduced a new specification test which can be used to compensate for the lack of robustness in the RE approach.

We conducted different simulations to study the small sample performance of FOSLS and FOSLS1 compared to the RML, first-differenced GMM, and the MPML. We considered four main scenarios and various levels of N, T and α_0. The results showed the competitive performance of the FOSLS whether the process is stationary or not and whether the error components are normally distributed or not. The outperformance of the FOSLS appears clearly when the distribution of the disturbances is skewed. The median absolute deviation of the RML was between 1.3 to 1.6 times higher than the median absolute deviation of the FOSLS under the nonnormal nonstationary scenario. However, the performance of the FOSLS1 wasn't stable for $N < T(T+3)/2$, so we don't recommend it for small N and large T panel data. Moreover, extra simulations were done to show that the good performance of the FOSLS is not affected if either a strictly exogenous variable was added to the structural autoregressive equation or the autoregressive parameter become very close to one. It turns out that the asymptotic formula of the FOSLS variance provides very accurate estimates of its small sample standard error.

Finally, we examined the practical usefulness of the SLS approach through an empirical study on the U.S. airfares data. We make use of all the developed tools in this chapter to find the best model which describes the determinants of airfares. Few messages are got from this empirical study. First, using the extra information inherent in the third and fourth moments of the process can refine the value of the estimates considerably and this may save important explanatory variables from being eliminated. Second, the FE approach may lead to unreliable estimates if there is a possibility to have weak instruments for the differenced time varying covariates. Last but not least, the level data is likely more informative than difference data, and this can improve the estimation precision and the goodness of fit considerably.

Conclusion and Future Work

In this dissertation we succeeded to generalize the SLS estimation approach of Wang (2003, 2004) in two study areas of dynamic modelling; the ARCH nonlinear regression model, and the linear dynamic panel data model. It is the first time to provide a rigorous theory for the SLS approach in dynamic framework. Specifically, in chapter 1 we generalized the SLS approach from the framework of cross sectional data to the framework of time series, and in chapter 2 we generalized the treatment of the SLS approach from static longitudinal data modelling to dynamic one. In both cases the results come encouraging. For the first time, we proposed a specification test for the SLS approach which can be used a diagnostic tool when we deal with real data application.

There are several possible generalizations and extensions of the SLS approach worth future investigation. For example, it is interesting to consider more general linear or nonlinear GARCH structure for the error variance. Generalization to multivariate processes should also be studied. Furthermore, since the SLS was originally proposed in the literature to handle the problem of measurement error in regression models, it is of a great interest to study the potential of using the SLS approach to treat the measurement error problem in the GARCH nonlinear regression models, and the nonlinear dynamic panel data models.

Appendices

Appendix A

Mathematical Proofs

Definition 6 *Let $\{\boldsymbol{Z}_t\}$ be a sequence of random vectors defined on a complete probability space $(\Omega, \mathcal{F}, P)$. Define $\mathcal{F}_{-\infty}^{n} = \mathcal{F}\left\{\boldsymbol{Z}_t,\, t \leq n\right\}$, $\mathcal{F}_{n+m}^{\infty} = \mathcal{F}\left\{\boldsymbol{Z}_t,\, t \geq n+m\right\}$, and*

$$\alpha(m) = \sup_{n} \sup_{\left\{F \in \mathcal{F}_{-\infty}^{n},\, G \in \mathcal{F}_{n+m}^{\infty}\right\}} \left|P(F \cap G) - P(F)P(G)\right|,$$

then $\{\boldsymbol{Z}_t\}$ is strong mixing of size $-a$ if $\alpha(m) = O(m^{-a-\delta})$, for some $\delta > 0$.

Proof of Theorem 1

The proof is done in four steps as follow.

Step 1. We apply the mean value theorem for random functions, as given in (Jennrich, 1969, Lemma 3), on the first order condition for a minimum of $Q_T(\boldsymbol{\gamma})$. Since $\boldsymbol{\gamma}_0$ is interior to Γ, there is a neighbourhood $\mathcal{N} \subset \Gamma$ of $\boldsymbol{\gamma}_0$. Let $F \in \mathcal{F}$ be the set with $P(F) = 1$ on which $\hat{\boldsymbol{\gamma}}_T(\omega) \rightarrow \boldsymbol{\gamma}_0$ as $T \rightarrow \infty$, $\hat{\boldsymbol{\gamma}}_T(\omega) = \operatorname{argmin}_{\boldsymbol{\gamma} \in \Gamma} Q_T(\omega, \boldsymbol{\gamma})$, and $f_T(\omega, \boldsymbol{\theta})$ is continuously differentiable of order 2 on Θ, uniformly in T, $T = 1, 2, \ldots$. By Jennrich's Lemma, there exist $\ddot{\boldsymbol{\gamma}}_T^i : \Omega \to \Gamma$ measurable–$\mathcal{F}_T$, $T = 1, 2, \ldots$ and $i = 1, 2, \ldots, (q+p+1)$, such that for $\omega \in F$ there exist $T_1(\omega) \in \mathbb{N}$, where $\hat{\boldsymbol{\gamma}}_T(\omega)$ is interior to $\mathcal{N}$, $\ddot{\boldsymbol{\gamma}}_T^i(\omega)$ is lying on the segment connecting $\hat{\boldsymbol{\gamma}}_T(\omega)$ and $\boldsymbol{\gamma}_0$, and $\nabla^2_{\boldsymbol{\gamma}} \ddot{Q}_T(\omega)\left(\hat{\boldsymbol{\gamma}}_T(\omega) - \boldsymbol{\gamma}_0\right) = -\nabla_{\boldsymbol{\gamma}} Q_T(\omega, \boldsymbol{\gamma}_0)$ for all $T > T_1(\omega)$, given that $\nabla^2_{\boldsymbol{\gamma}} \ddot{Q}_T(\omega)$ is the $(q+p+1) \times (q+p+1)$ Hessian matrix $\nabla^2_{\boldsymbol{\gamma}} Q_T(\omega, \boldsymbol{\gamma})$ with i th row evaluated at $\ddot{\boldsymbol{\gamma}}_T^i(\omega)$.

Step 2. Let $\bar{\boldsymbol{A}}_T(\boldsymbol{\gamma}) = 2T^{-1}\sum_{t=1}^{T} E\left\{\nabla_{\boldsymbol{\gamma}} \boldsymbol{h}_t'(\boldsymbol{\gamma}) \boldsymbol{W}_t \nabla_{\boldsymbol{\gamma}'} \boldsymbol{h}_t(\boldsymbol{\gamma})\right\}$, then we show there exist $F' \subset F$ with $P(F') = 1$ such that for $\omega \in F'$, we have

$$\left\| \nabla^2_{\boldsymbol{\gamma}} \ddot{Q}_T(\omega) - \bar{\boldsymbol{A}}_T(\boldsymbol{\gamma}_0) \right\| \longrightarrow 0 \text{ as } T \longrightarrow \infty. \tag{A.1}$$

By Hölder's inequality, triangle inequality, and c_r inequality, it turns out that assumptions 2, 6 are sufficient to verify that $\|\boldsymbol{A}_t(\boldsymbol{\gamma})\|$ are dominated by uniformly L_r-bounded functions, where $\boldsymbol{A}_t(\boldsymbol{\gamma}) = 2\nabla_{\boldsymbol{\gamma}} \boldsymbol{h}_t'(\boldsymbol{\gamma}) \boldsymbol{W}_t \nabla_{\boldsymbol{\gamma}'} \boldsymbol{h}_t(\boldsymbol{\gamma}) + 2\left(\boldsymbol{h}_t'(\boldsymbol{\gamma}) \boldsymbol{W}_t \bigotimes \boldsymbol{I}_{q+p+1}\right) \nabla_{\boldsymbol{\gamma}'} vec\left(\nabla_{\boldsymbol{\gamma}} \boldsymbol{h}_t'(\boldsymbol{\gamma})\right)$. Hence, by the the uniform law of large numbers given in (White and Domowitz, 1984, Theorem 2.3) we have for $\omega \in F'$

$$\sup_{\boldsymbol{\gamma} \in \Gamma} \left\| \nabla^2_{\boldsymbol{\gamma}} Q_T(\omega, \boldsymbol{\gamma}) - T^{-1} \sum_{t=1}^{T} E \boldsymbol{A}_t(\boldsymbol{\gamma}) \right\| \to 0 \text{ as } T \longrightarrow \infty.$$

Moreover, by the triangle inequality we have

$$\left\|\nabla_{\boldsymbol{\gamma}}^2 \ddot{Q}_T(\omega) - T^{-1}\sum_{t=1}^{T} E\boldsymbol{A}_t(\boldsymbol{\gamma}_0)\right\| \leq \sup_{\boldsymbol{\gamma}\in\Gamma}\left\|\nabla_{\boldsymbol{\gamma}}^2 Q_T(\omega,\boldsymbol{\gamma}) - T^{-1}\sum_{t=1}^{T} E\boldsymbol{A}_t(\boldsymbol{\gamma})\right\|$$

$$+ \sup_{K\in\mathcal{N}}\left\|K^{-1}\sum_{t=1}^{K}\ddot{\boldsymbol{A}}_{t:T}(\omega) - K^{-1}\sum_{t=1}^{K} E\boldsymbol{A}_t(\boldsymbol{\gamma}_0)\right\| \quad \forall\,\omega\in F',$$

where $\ddot{\boldsymbol{A}}_{t:T}(\omega)$ is $E\boldsymbol{A}_t(\boldsymbol{\gamma})$ with i th row evaluated at $\ddot{\boldsymbol{\gamma}}_T^i(\omega)$. Since $\ddot{\boldsymbol{\gamma}}_T^i(\omega)\to\boldsymbol{\gamma}_0$ as $T\to\infty$, $E\boldsymbol{A}_t(\boldsymbol{\gamma})$ is continuous on Γ uniformly in t, and $E\boldsymbol{A}_t(\boldsymbol{\gamma}_0) = 2E\left\{\nabla_{\boldsymbol{\gamma}}\boldsymbol{h}_t'(\boldsymbol{\gamma}_0)\boldsymbol{W}_t\nabla_{\boldsymbol{\gamma}'}\boldsymbol{h}_t(\boldsymbol{\gamma}_0)\right\}$, then equation (A.1) is obtained immediately by letting $T\to\infty$ in the last inequality. Since $\bar{\boldsymbol{A}}_T(\boldsymbol{\gamma}_0)$ and $\boldsymbol{V}_T$ are uniformly nonsingular for all T sufficiently large (assumptions 7 and 9 respectively), then for $\omega\in F'$ there exist $T_2(\omega)\in\mathbb{N}$ such that for all $T>T_2(\omega)$,

$$\boldsymbol{V}_T^{-1/2}\bar{\boldsymbol{A}}_T(\boldsymbol{\gamma}_0)\sqrt{T}\left(\hat{\boldsymbol{\gamma}}_T(\omega)-\boldsymbol{\gamma}_0\right) = -\boldsymbol{V}_T^{-1/2}\sqrt{T}\nabla_{\boldsymbol{\gamma}}Q_T(\omega,\boldsymbol{\gamma}_0)$$

$$+\boldsymbol{V}_T^{-1/2}\bar{\boldsymbol{A}}_T(\boldsymbol{\gamma}_0)\left(\bar{\boldsymbol{A}}_T^{-1}(\boldsymbol{\gamma}_0)-\nabla_{\boldsymbol{\gamma}}^2\ddot{Q}_T^{-1}(\omega)\right)\boldsymbol{V}_T^{1/2}\boldsymbol{V}_T^{-1/2}\sqrt{T}\nabla_{\boldsymbol{\gamma}}Q_T(\omega,\boldsymbol{\gamma}_0). \tag{A.2}$$

Step 3. We use Cramér-Wold device (Rao, 1973, p. 123) to show that

$\boldsymbol{V}_T^{-1/2}\sqrt{T}\nabla_{\boldsymbol{\gamma}}Q_T(\boldsymbol{\gamma}_0)\xrightarrow{d} N(\boldsymbol{0},\boldsymbol{I}_{q+p+1})$ as $T\to\infty$. Let $\boldsymbol{\lambda}\in\mathbb{R}^{q+p+1}$ such that $\|\boldsymbol{\lambda}\|=1$, then it is enough to show that $T^{-1/2}\sum_{t=1}^{T}\boldsymbol{\lambda}'\boldsymbol{V}_T^{-1/2}\boldsymbol{S}_t(\boldsymbol{\gamma}_0)\xrightarrow{d} N(0,1)$ as $T\to\infty$, where $\boldsymbol{S}_t(\boldsymbol{\gamma}_0)=2\nabla_{\boldsymbol{\gamma}}\boldsymbol{h}_t'(\boldsymbol{\gamma}_0)\boldsymbol{W}_t\boldsymbol{h}_t(\boldsymbol{\gamma}_0)$. By assumption 9 we have $\boldsymbol{V}_T^{-1/2}=\boldsymbol{O}(1)$ and therefore it turns out by using Hölder's inequality and c_r inequality that assumption 8 is sufficient for the double array of scalars $\left\{m_{Tt}=\boldsymbol{\lambda}'\boldsymbol{V}_T^{-1/2}\boldsymbol{S}_t(\boldsymbol{\gamma}_0)\right\}$ to be uniformly L_r-bounded for all T sufficiently large. Since $\{\boldsymbol{h}_t(\boldsymbol{\gamma}_0), \mathcal{F}_t\}$ is a martingale difference, then $E\left(m_{Tt}\right) = 0$, and $\text{var}\left(T^{-1/2}\sum_{t=1}^{T} m_{Tt}\right) = 1$ for all T sufficiently large. By Theorem 14.1 of Davidson (1994), assumption 1 is sufficient for $\{m_{Tt}\}$ to be strong mixing of size $-a$, and hence by Theorem 5.20 of White (2001) we have $T^{-1/2}\sum_{t=1}^{T} m_{Tt} \xrightarrow{d} N\left(0,1\right)$ as $T \to \infty$.

Step 4. Viewing $\ddot{\boldsymbol{\gamma}}_T^i$ as pointwise limits of sequences of simple functions, then $\nabla_{\boldsymbol{\gamma}}^2 \ddot{Q}_T$ are measurable, and from step 2 we have $\left\|\nabla_{\boldsymbol{\gamma}}^2 \ddot{Q}_T - \bar{\boldsymbol{A}}_T(\boldsymbol{\gamma}_0)\right\| \xrightarrow{a.s.} 0$ as $T \to \infty$. By assumption 7 and Proposition 2.16 of White (2001), we can show that $\left(\nabla_{\boldsymbol{\gamma}}^2 \ddot{Q}_T^{-1} - \bar{\boldsymbol{A}}_T^{-1}(\boldsymbol{\gamma}_0)\right)$ is $\boldsymbol{o}_p(1)$. Since $\boldsymbol{V}_T^{-1/2}\sqrt{T}\nabla_{\boldsymbol{\gamma}} Q_T\left(\boldsymbol{\gamma}_0\right)$ is $\boldsymbol{O}_p(1)$ from step 3 then it follows that

$$\boldsymbol{V}_T^{-1/2}\bar{\boldsymbol{A}}_T(\boldsymbol{\gamma}_0)\left(\bar{\boldsymbol{A}}_T^{-1}(\boldsymbol{\gamma}_0) - \nabla_{\boldsymbol{\gamma}}^2 \ddot{Q}_T^{-1}\right)\sqrt{T}\nabla_{\boldsymbol{\gamma}} Q_T\left(\boldsymbol{\gamma}_0\right) = \boldsymbol{o}_p(1).$$

By applying the method of subsequences (Davidson, 1994, Theorem 18.6) to a subsequence indexed by $\{T'\}$, there exists a further subsequence indexed by $\{T''\}$ such that

$$\boldsymbol{V}_{T''}^{-1/2}\bar{\boldsymbol{A}}_{T''}(\boldsymbol{\gamma}_0)\left(\bar{\boldsymbol{A}}_{T''}^{-1}(\boldsymbol{\gamma}_0) - \nabla_{\boldsymbol{\gamma}}^2 \ddot{Q}_{T''}^{-1}\right)\sqrt{T''}\nabla_{\boldsymbol{\gamma}} Q_{T''}\left(\boldsymbol{\gamma}_0\right) \xrightarrow{a.s.} 0 \text{ as } T'' \longrightarrow \infty,$$

and therefore

$$\boldsymbol{V}_{T''}^{-1/2}\sqrt{T''}\left(\bar{\boldsymbol{A}}_{T''}(\boldsymbol{\gamma}_0)\left(\hat{\boldsymbol{\gamma}}_{T''}\left(\omega\right) - \boldsymbol{\gamma}_0\right) + \nabla_{\boldsymbol{\gamma}} Q_{T''}\left(\omega, \boldsymbol{\gamma}_0\right)\right) \longrightarrow 0 \quad \text{a.s.-}P \text{ as } T'' \longrightarrow \infty.$$

Since $\{T'\}$ is arbitrarily chosen, then

$$\boldsymbol{V}_T^{-1/2}\bar{\boldsymbol{A}}_T(\boldsymbol{\gamma}_0)\sqrt{T}\left(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}_0\right) + \boldsymbol{V}_T^{-1/2}\sqrt{T}\nabla_{\boldsymbol{\gamma}} Q_T\left(\boldsymbol{\gamma}_0\right) \longrightarrow 0 \quad \text{prob-}P \text{ as } T \longrightarrow \infty,$$

and the proof is completed by applying result (2c.4.12) of Rao (1973).

□

Proof of proposition 2

The proof follows by noting that for all $\boldsymbol{a} \in \mathbb{R}^{q+p+1}$,

$$\boldsymbol{a}' E\left\{\left(\boldsymbol{R} - \boldsymbol{M}\, E^{-1}\left\{\boldsymbol{M}'\boldsymbol{M}\right\} E\left\{\boldsymbol{M}'\boldsymbol{R}\right\}\right)'\left(\boldsymbol{R} - \boldsymbol{M}\, E^{-1}\left\{\boldsymbol{M}'\boldsymbol{M}\right\} E\left\{\boldsymbol{M}'\boldsymbol{R}\right\}\right)\right\}\boldsymbol{a} \geq 0, \tag{A.3}$$

where

$\boldsymbol{R}' = \frac{1}{\sqrt{T}}\left(\boldsymbol{R}_1', \boldsymbol{R}_2', \ldots, \boldsymbol{R}_T'\right)$, $\boldsymbol{M}' = \frac{1}{\sqrt{T}}\left(\boldsymbol{M}_1', \boldsymbol{M}_2', \ldots, \boldsymbol{M}_T'\right)$, $\boldsymbol{R}_t' = \nabla_{\boldsymbol{\gamma}}\boldsymbol{h}_t'(\boldsymbol{\gamma}_0)\boldsymbol{W}_t\boldsymbol{U}_t^{1/2}$, and $\boldsymbol{M}_t' = \nabla_{\boldsymbol{\gamma}}\boldsymbol{h}_t'(\boldsymbol{\gamma}_0)\boldsymbol{U}_t^{-1/2}$. Moreover the equality in (A.3) holds if $\boldsymbol{W}_t = \boldsymbol{U}_t^{-1}$, $t = 1, 2, \ldots, T$, which justifies equation (1.4). The equivalence between equations (1.4) and (1.5) follows form substituting $\boldsymbol{\Omega}_t^{-1}$ and $\boldsymbol{B}_t$ in equation (1.5). □

Proof of proposition 3

For the simplicity of notation, we use f_j, y_t, $\tilde{\boldsymbol{x}}_t$, a_t, c_{ts}, d_{ts}, $\boldsymbol{w}_{ts}$, $\boldsymbol{k}_{ts}$ for $f_j(y_{10}, \boldsymbol{x}_1, \boldsymbol{\theta})$, y_{1t}, $\tilde{\boldsymbol{x}}_{1t}(\alpha)$, $a_t(\alpha)$, $c_{ts}(\alpha)$, $d_{ts}(\alpha)$, $\boldsymbol{w}_{1ts}(\alpha)$, and $\boldsymbol{k}_{1ts}(\alpha)$ respectively.

First, by Cauchy-Schwarz inequality we have

$$
\begin{aligned}
\|\boldsymbol{h}_1(\boldsymbol{\gamma})\|^2 \leq {} & 2\sum_{t=1}^{T} y_t^2 + 8y_0^2 \sum_{t=1}^{T} \alpha^{2t} + 8f_1^2 \sum_{t=1}^{T} a_t^2 + 4\sum_{t=1}^{T} \boldsymbol{\beta}'\tilde{\boldsymbol{x}}_t\tilde{\boldsymbol{x}}_t'\boldsymbol{\beta} + 2\sum_{1\leq s\leq t} y_t^2 y_s^2 \\
& + 16y_0^4 \sum_{1\leq s\leq t}^{T} \alpha^{2(t+s)} + 16f_2^2 \sum_{1\leq s\leq t}^{T} a_t^2 a_s^2 + 16\sigma^4 c_{ts}^2 + 16y_0^2 f_1^2 \sum_{1\leq s\leq t}^{T} d_{ts}^2 \\
& + 8\sum_{1\leq s\leq t}^{T} (\boldsymbol{\beta}'\tilde{\boldsymbol{x}}_t\tilde{\boldsymbol{x}}_s'\boldsymbol{\beta})^2 + 16y_0^2 \sum_{1\leq s\leq t}^{T} (\boldsymbol{\beta}'\boldsymbol{w}_{ts})^2 + 16f_1^2 \sum_{1\leq s\leq t}^{T} (\boldsymbol{\beta}'\boldsymbol{k}_{ts})^2,
\end{aligned}
$$

and

$$
\begin{aligned}
E\|\boldsymbol{W}_1\|(y_t y_s)^2 &\leq E\|\boldsymbol{W}_1\| y_t^4 \, E\|\boldsymbol{W}_1\| y_s^4, \\
E\|\boldsymbol{W}_1\| y_0^2 f_1^2 &\leq E\|\boldsymbol{W}_1\| y_0^4 \, E\|\boldsymbol{W}_1\| f_2^2.
\end{aligned}
$$

Therefore by assumptions 13 and 14 we have

$$
E\sup_{\boldsymbol{\Gamma}} |q_1(\boldsymbol{\gamma})| \leq E\|\boldsymbol{W}_1\| \sup_{\boldsymbol{\Gamma}} \|\boldsymbol{h}_1(\boldsymbol{\gamma})\|^2 < \infty.
$$

It follows from the uniform law of large numbers (ULLN Jennrich, 1969; Amemiya, 1985) that

$$
\sup_{\boldsymbol{\Gamma}} \left| \frac{1}{N}\sum_{i=1}^{N} q_i(\boldsymbol{\gamma}) - Eq_1(\boldsymbol{\gamma}) \right| \xrightarrow{a.s.} 0 \quad \text{as } N \to \infty \text{ and } T \text{ is fixed.} \tag{A.4}
$$

Second, since

$$
\begin{aligned}
Eq_1(\boldsymbol{\gamma}) = {} & E\boldsymbol{h}_1'(\boldsymbol{\gamma}_0) W_1 \boldsymbol{h}_1(\boldsymbol{\gamma}_0) + 2E\left[\boldsymbol{h}_1(\boldsymbol{\gamma}) - \boldsymbol{h}_1(\boldsymbol{\gamma}_0)\right]' W_1 \boldsymbol{h}_1(\boldsymbol{\gamma}_0) \\
& + E\left[\boldsymbol{h}_1(\boldsymbol{\gamma}) - \boldsymbol{h}_1(\boldsymbol{\gamma}_0)\right]' W_1 \left[\boldsymbol{h}_1(\boldsymbol{\gamma}) - \boldsymbol{h}_1(\boldsymbol{\gamma}_0)\right]
\end{aligned}
$$

and the second term is apparently equal to zero, $Eq_1(\boldsymbol{\gamma}) \geq Eq_1(\boldsymbol{\gamma}_0)$ and the equality holds if and only if $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$. Finally the result follows from (A.4) and Lemma 1 of Wang and Leblanc (2008). □

Proof of proposition 4

For the simplicity of notation, we use f_j, y_t, $\tilde{\boldsymbol{x}}_t$, a_t, c_{ts}, d_{ts}, $\boldsymbol{w}_{ts}$, $\boldsymbol{k}_{ts}$ for $f_j(y_{10}, \boldsymbol{x}_1, \boldsymbol{\theta})$, y_{1t}, $\tilde{\boldsymbol{x}}_{1t}(\alpha)$, $a_t(\alpha)$, $c_{ts}(\alpha)$, $d_{ts}(\alpha)$, $\boldsymbol{w}_{1ts}(\alpha)$, and $\boldsymbol{k}_{1ts}(\alpha)$ respectively.

First, by the mean value theorem for random functions (Jennrich, 1969), assumptions 13–16 guarantee that

$$1_{\hat{\boldsymbol{\gamma}}_N \in \mathcal{N}(\boldsymbol{\gamma}_0)} \left[\sum_{i=1}^{N} \boldsymbol{s}_i(\boldsymbol{\gamma}_0) + \left(\sum_{i=1}^{N} \bar{\boldsymbol{H}}_i \right) (\hat{\boldsymbol{\gamma}}_N - \boldsymbol{\gamma}_0) \right] = \boldsymbol{0}, \tag{A.5}$$

where $\boldsymbol{s}_i(\boldsymbol{\gamma}) = \frac{\partial q_i(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = 2\frac{\partial \boldsymbol{h}_i'(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \boldsymbol{W}_i \boldsymbol{h}_i(\boldsymbol{\gamma})$, and the rth row of $\bar{\boldsymbol{H}}_i$ is given by

$$\frac{\partial^2 q_i(\bar{\boldsymbol{\gamma}}_N^r)}{\partial \boldsymbol{\gamma}_{(r)} \partial \boldsymbol{\gamma}'}, \quad r = 1, 2, \ldots, (p + \ell + 2),$$

and $\bar{\boldsymbol{\gamma}}_N^r$ are measurable mappings into $\mathcal{N}(\boldsymbol{\gamma}_0)$ and they lie on the segment joining $\hat{\boldsymbol{\gamma}}_N$ and $\boldsymbol{\gamma}_0$.

Next we use the consistency of $\hat{\boldsymbol{\gamma}}_N$ and the ULLN to show that, for fixed T,

$$\frac{1}{N} \sum_{i=1}^{N} \bar{\boldsymbol{H}}_i \xrightarrow{a.s.} 2\boldsymbol{A} \quad \text{as} \quad N \to \infty.$$

Starting by the triangle inequality we have

$$\left\| \frac{\partial^2 q_1(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} \right\| \leq 2 \left\| \frac{\partial \boldsymbol{h}_1'(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \boldsymbol{W}_1 \frac{\partial \boldsymbol{h}_1(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}'} \right\| + 2 \left\| \left(\frac{\partial^2 \boldsymbol{h}_1'(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_{(i)} \partial \boldsymbol{\gamma}_{(j)}} \boldsymbol{W}_1 \boldsymbol{h}_1(\boldsymbol{\gamma}) \right)_{i,j} \right\|, \tag{A.6}$$

and further by Cauchy-Schwarz inequality

$$E \sup_{\mathcal{N}(\boldsymbol{\gamma}_0)} \left\| \frac{\partial \boldsymbol{h}_1'(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \boldsymbol{W}_1 \frac{\partial \boldsymbol{h}_1(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}'} \right\| \leq E \|\boldsymbol{W}_1\| \sup_{\mathcal{N}(\boldsymbol{\gamma}_0)} \left\| \frac{\partial \boldsymbol{h}_1'(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \right\|^2,$$

and the following inequalities hold for $1 \leq s \leq t \leq T$,

$$\left(\frac{\partial \mu_{1t}(\boldsymbol{\gamma})}{\partial \alpha} \right)^2 \leq 4t^2 \alpha^{2(t-1)} y_0^2 + 4 \left(\frac{\partial a_t}{\partial \alpha} \right)^2 f_2 + 2 \left(\boldsymbol{\beta}' \frac{\partial \tilde{\boldsymbol{x}}_t}{\partial \alpha} \right)^2,$$

$$\begin{aligned} \left(\frac{\partial \nu_{1ts}(\boldsymbol{\gamma})}{\partial \alpha} \right)^2 \leq{}& 8(t+s)^2 \alpha^{2(t+s-1)} y_0^4 + 8 f_2^2 \left(a_s \frac{\partial a_t}{\partial \alpha} + a_t \frac{\partial a_s}{\partial \alpha} \right)^2 + 8\sigma^4 \left(\frac{\partial c_{ts}}{\partial \alpha} \right)^2 \\ &+ 8 y_0^2 f_2 \left(\frac{\partial d_{ts}}{\partial \alpha} \right)^2 + 8 y_0^2 \left(\boldsymbol{\beta}' \frac{\partial \boldsymbol{w}_{ts}}{\partial \alpha} \right)^2 + 8 f_2 \left(\boldsymbol{\beta}' \frac{\partial \boldsymbol{k}_{ts}}{\partial \alpha} \right)^2 \\ &+ 8 \left(\boldsymbol{\beta}' \frac{\partial \tilde{\boldsymbol{x}}_t}{\partial \alpha} \right)^2 (\boldsymbol{\beta}' \tilde{\boldsymbol{x}}_s)^2 + 8 \left(\boldsymbol{\beta}' \frac{\partial \tilde{\boldsymbol{x}}_s}{\partial \alpha} \right)^2 (\boldsymbol{\beta}' \tilde{\boldsymbol{x}}_t)^2, \end{aligned}$$

$$\left\| \frac{\partial \nu_{1ts}(\boldsymbol{\gamma})}{\partial \boldsymbol{\theta}} \right\|^2 \leq 4 a_t^2 a_s^2 \left\| \frac{\partial f_2}{\partial \boldsymbol{\theta}} \right\|^2 + 4 d_{ts}^2 y_0^2 \left\| \frac{\partial f_1}{\partial \boldsymbol{\theta}} \right\|^2 + 2 \left(\boldsymbol{\beta}' \boldsymbol{k}_{ts} \right)^2 \left\| \frac{\partial f_1}{\partial \boldsymbol{\theta}} \right\|^2,$$

$$\left\| \frac{\partial \nu_{1ts}(\boldsymbol{\gamma})}{\partial \boldsymbol{\beta}} \right\|^2 \leq 4 \left\| \tilde{\boldsymbol{x}}_t \tilde{\boldsymbol{x}}_s' \right\|^2 \|\boldsymbol{\beta}\|^2 + 4 y_0^2 \|\boldsymbol{w}_{ts}\|^2 + 4 f_2 \|\boldsymbol{k}_{ts}\|^2.$$

Then, by assumptions 14 and 16, it is clear that

$$E \|\boldsymbol{W}_1\| \sup_{\mathcal{N}(\boldsymbol{\gamma}_0)} \left\| \frac{\partial \boldsymbol{h}_1'(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} \right\|^2 < \infty. \tag{A.7}$$

With similar argument, by the Cauchy-Schwarz inequality and assumptions 14 and 16 we have

$$\begin{aligned} E \sup_{\mathcal{N}(\boldsymbol{\gamma}_0)} \left\| \left(\frac{\partial^2 \boldsymbol{h}_1'(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_{(i)} \partial \boldsymbol{\gamma}_{(j)}} \boldsymbol{W}_1 \boldsymbol{h}_1(\boldsymbol{\gamma}) \right)_{i,j} \right\| &\leq E\|\boldsymbol{W}_1\| \sup_{\mathcal{N}(\boldsymbol{\gamma}_0)} \|\boldsymbol{h}_1(\boldsymbol{\gamma})\|^2 \\ &\quad \times E\|\boldsymbol{W}_1\| \sup_{\mathcal{N}(\boldsymbol{\gamma}_0)} \left(\sum_{i,j}^{p+\ell+2} \left\| \frac{\partial^2 \boldsymbol{h}_1'(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_{(i)} \partial \boldsymbol{\gamma}_{(j)}} \right\|^2 \right) \\ &< \infty. \end{aligned} \tag{A.8}$$

By combining inequalities (A.6, A.7, and A.8) together with assumptions 14 and 16, the ULLN implies that

$$\sup_{\mathcal{N}(\boldsymbol{\gamma}_0)} \left\| \frac{1}{N} \sum_{i=1}^{N} \frac{\partial^2 q_i(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} - E \frac{\partial^2 q_1(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} \right\| \xrightarrow{a.s.} 0 \quad \text{as} \quad N \to \infty \quad \text{for fixed } T. \tag{A.9}$$

According to Lemma 2 of Wang and Leblanc (2008), the strong consistency of $\hat{\boldsymbol{\gamma}}_N$ and equation (A.9) are sufficient to conclude that

$$\frac{1}{N} \sum_{i=1}^{N} \bar{\boldsymbol{H}}_i \xrightarrow{a.s.} E \frac{\partial^2 q_1(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} = 2\boldsymbol{A} \quad \text{as} \quad N \to \infty \quad \text{for fixed } T. \tag{A.10}$$

Finally, under assumption 18 the multivariate central limit theorem implies that

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{s}_i(\boldsymbol{\gamma}_0) \xrightarrow{d} N(\boldsymbol{0}, 4\boldsymbol{B}) \quad \text{as} \quad N \to \infty \quad \text{for fixed } T,$$

where $\boldsymbol{B}$ is given in Theorem 4. Hence by Slutzky theorem and equations (A.5, A.10), we obtain $\sqrt{N}(\hat{\boldsymbol{\gamma}}_N - \boldsymbol{\gamma}_0) = -(2\boldsymbol{A})^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \boldsymbol{s}_i(\boldsymbol{\gamma}_0) + \boldsymbol{o}_p(1)$, for fixed T. $\square$

The asymptotic variance of the RMLE

Using the GMM framework we derive the asymptotic variance of the RML estimator under model (2.1). It is given by

$$\boldsymbol{F}_0 = E^{-1}\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right) E\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{M}\boldsymbol{V}\boldsymbol{K}\right) E^{-1}\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right), \tag{A.11}$$

where

$$\boldsymbol{K}' = \begin{pmatrix} \dfrac{\partial \boldsymbol{\mu}'(\boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}} & \dfrac{\partial vech'(\boldsymbol{S}(\boldsymbol{\gamma}_0))}{\partial \boldsymbol{\gamma}} \end{pmatrix}, \quad \boldsymbol{\mu}'(\boldsymbol{\gamma}) = \left(\mu_{1t}(\boldsymbol{\gamma}), 1 \le t \le T\right),$$

$$vech'(\boldsymbol{S}(\boldsymbol{\gamma})) = \left(a_t a_s e^{\theta_3} + \sigma^2 c_{ts},\ 1 \le s \le t \le T\right),$$

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{S}(\boldsymbol{\gamma}_0)^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{1}{2}\boldsymbol{L}'\left[\boldsymbol{S}(\boldsymbol{\gamma}_0)^{-1} \otimes \boldsymbol{S}(\boldsymbol{\gamma}_0)^{-1}\right]\boldsymbol{L} \end{pmatrix}, \quad vec(\boldsymbol{S}(\boldsymbol{\gamma})) = \boldsymbol{L}\, vech(\boldsymbol{S}(\boldsymbol{\gamma})),$$

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{S}(\boldsymbol{\gamma}_0) & E\left\{\boldsymbol{u}_1 vech'(\boldsymbol{u}_1\boldsymbol{u}_1')|y_{10}, \boldsymbol{x}_1\right\} \\ . & E\left\{vech(\boldsymbol{u}_1\boldsymbol{u}_1')vech'(\boldsymbol{u}_1\boldsymbol{u}_1')|y_{10}, \boldsymbol{x}_1\right\} - vech(\boldsymbol{S}(\boldsymbol{\gamma}_0))vech'(\boldsymbol{S}(\boldsymbol{\gamma}_0)) \end{pmatrix},$$

$$\boldsymbol{u}_1 = \left(y_{1t} - \mu_{1t}(\boldsymbol{\gamma}_0), 1 \le t \le T\right)'.$$

By using the same notations, the asymptotic variance of the optimal SLS estimator is given by

$$\boldsymbol{F}_0^* = E^{-1}\left(\boldsymbol{K}'\boldsymbol{M}^{-1}\boldsymbol{K}\right). \tag{A.12}$$

It is easy to show that $\boldsymbol{F}_0 - \boldsymbol{F}_0^*$ is nonnegative definite and the difference vanishes if and only if

$$\boldsymbol{M}\boldsymbol{V}\boldsymbol{K} = \boldsymbol{K}E^{-1}\left(\boldsymbol{K}'\boldsymbol{M}^{-1}\boldsymbol{K}\right) E\left(\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K}\right). \tag{A.13}$$

The *vec* operator stacks by columns a matrix, and the *vech* operator stacks by column the lower triangle of a square matrix.

Appendix B

Sample of R Programming Code

```
##########################################
#The following code was used to produced The results of tables B11-B12
##########################################

rm(list=ls())
local({r <- getOption("repos")
       r["CRAN"] <- "http://cran.stat.sfu.ca/"
       options(repos=r)})
pkg_list = c('snow', 'snowfall','MASS','matrixcalc','plm','nlme','Matrix')
for (pkg in pkg_list)
{
  # Try loading the library.
  if ( ! library(pkg, logical.return=TRUE, character.only=TRUE) )
  {
    # If the library cannot be loaded, install it; then load.
    install.packages(pkg,dependencies=T)
    library(pkg, character.only=TRUE)
```

```
  }
}
#Define the model parameters (exactly like kivit 1995)

#fixed values
n=100;vareps=1

#variable values
mu=1;pro=0.8;alpha=0.4;t=3;sigm2_s=2


#marginal variance of eta
vareta=(1-alpha)^2*vareps*mu^2
beta=1-alpha
sigm2_exi=(sigm2_s-alpha^2/(1-alpha^2)*vareps)*
(1+(alpha+pro)^2/(1+alpha*pro)*
(alpha*pro-1)-(alpha*pro)^2)/beta^2
p1=(alpha+pro)/(1+alpha*pro)
p2=(alpha+pro)^2/(1+alpha*pro)-alpha*pro
varphi_0=sigm2_exi/(1-(alpha+pro)*p1+alpha*pro*p2)
varepsi_0=vareps/(1-alpha^2)
vary0=varphi_0*beta^2+varepsi_0+vareta/(1-alpha)^2
covy0eta=vareta/(1-alpha)
varx1=sigm2_exi/(1-pro^2)
theta0=log(vareps)

# the slope of (eta on y0 equation)
```

```
theta2=covy0eta/(vary0-beta^2*varphi_0*pro^2)

#the intercept of eta on y0,x1
theta1=0

# theta3 is reparametrized conditional variance of eta given y0,x1
theta3=log(vareta-covy0eta^2*varx1/(vary0*varx1-beta^2*varphi_0*
sigm2_exi*pro/(1-pro^2)))

#the solope of (eta on x1)
theta4=theta2*beta*pro*sqrt(varphi_0/sigm2_exi*(1-pro^2))


#centeral marginal and conditional third moment of epsilon (disturbance )
m3eps=0

#centeral  conditional third moment of eta on y0 an x1(subject effect)
m3eta=0

#centeral marginal and conditional fourth moment of epsilon
m4eps=3*vareps^2

#centeral conditional fourth moment of eta
m4eta=3*exp(2*theta3)

gamm0=c(alpha,beta,theta0,theta1,theta2,theta3,theta4)

#number fo simulation runs
ss=1000
```

```
#####################################################
#Define the GMM1 function and it's supplementary matrix
#(This is the GMM of Arelano baond 1991 simulation, and also kiviet 1995)
####################################################
#Difference Transformation
Dff=cbind(diag(rep(-1,t-1)),rep(0,t-1))+cbind(rep(0,t-1),diag(rep(1,t-1)))


GMM=function(v){
  y=v[1:(t+1)]
  x=v[(t+2):(2*(t+1))]
  Z=y[1]
  for (i in 1:(t-2)){
    Z=bdiag(Z,y[1:(i+1)])
  }
  Dffx=Dff%*%x[2:(t+1)]
  Z=rbind(as.matrix(Z),t(Dffx))
  Dffy=Dff%*%y[2:(t+1)]
  Dffy_1=Dff%*%y[1:(t)]
  weight=Z%*%Dff%*%t(Dff)%*%t(Z)
  return(c(Z%*%Dffx,Z%*%Dffy,Z%*%Dffy_1,as.vector(weight)))
}


#####RML related matrices########
#define the forward orthoganal opertator
```

```
FOO1=matrix(1,t-1,t-1)
FOO1[lower.tri(FOO1,diag=T)]=0
FOO=diag(((t-1):1)/(t:2))^.5%*%
  (diag(-1/((t-1):1))%*%cbind(FOO1,1)+cbind(diag(1,t-1),0))
######################################
#compute the projection matrix
proj<-function(X,n){
  Y=matrix(t(X),nrow=n)
  return(Y%*%solve(t(Y)%*%Y,t(Y)))

}
##################################################
#computing the RML (CGLS) using the concentrated likelihood as in Arelano
#book chapter 6
###############################################
cgls=function(param){
  alpha=param[1];beta=param[2]
  FOOY=FOO%*%B[2:(t+1),,1]
  FOOY_1=FOO%*%B[1:(t),,1]
  FOOX=FOO%*%B[2:(t+1),,2]
  Part1=log(sum(as.vector((FOOY-alpha*FOOY_1-beta*FOOX)^2)))
  ybar=apply(B[2:(t+1),,1],MARGIN=2,mean)
  y_1bar=apply(B[1:(t),,1],MARGIN=2,mean)
  xbar=apply(B[2:(t+1),,2],MARGIN=2,mean)
  Part2=1/(t-1)*log(sum(lm(ybar-alpha*y_1bar-beta*xbar~B[1,,1]+
  B[2,,2])$residuals^2))
```

```
  return(Part1+Part2)
}


#Auxilary matrices
plus=outer((1:t),(1:t),'+')
mins=outer((1:t),(1:t),'-')

#calculating the objective function of the SLS
fuu<-function(gam,W){
     a=gam[1]^(1:t)   #alpha^t
  fa<-cumsum(gam[1]^(0:(t-1))) #a_t
  FA<-gam[1]^mins
  fc<-FA*(rep(1,t)%o%cumsum(gam[1]^(2*(0:(t-1)))))
  fc[upper.tri(fc)]=t(fc)[upper.tri(fc)]#c_ts
  fd<-a%o%fa+fa%o%a #d_ts
  FA[upper.tri(FA)]=0
  FH=apply(rbind(B[,,1],B[,,2]),MARGIN=2,FUN=fh,gam,fa,fc,fd,a,FA)
    obj=0
  for (k in 1:n){
        obj=obj+ FH[,k]%*% matrix(W[,k],t*(t+3)/2,t*(t+3)/2)%*% FH[,k]
    }
  return(obj)
  }

##########the vector h in objective function of the SLS
```

```
fh<-function(yy,gam,fa,fc,fd,a,FA){
  y=yy[1:(t+1)]
  x=yy[-(1:(t+1))]
  f1<-gam[4]+y[1]*gam[5]+x[2]*gam[7]#f_1
  f2<-exp(gam[6])+f1^2#f_2
  fax=as.vector(FA%*%x[-1])#x_t^tilda
  wts<-a%o%fax+fax%o%a #w_ts
  kts<-fa%o%fax+fax%o%fa #d_ts
  h=c(
    y[-1]-a*y[1]-fa*f1-gam[2]*fax,
    (y[-1]%o%y[-1]-(gam[1]^plus*y[1]^2+
         fa%o%fa*f2+exp(gam[3])*fc+fd*y[1]*f1+fax%o%fax*gam[2]^2+
 wts*y[1]*gam[2]+kts*gam[2]*f1
                   ))[lower.tri(diag(t),diag=T)]
  )
  return(h)
}

#This  function is auxilary to  calculate  optimal weight matrix of SLS
Trans<-function(K,y,gam){
    x=y[-seq(K+2)]
  y0=y[1]
  f1<-gam[3]+y0*gam[4]+gam[7]*x[1]
    R=diag(nrow=K*(K+3)/2)
  for (k in 2:K)
```

```
{R[k,k-1]=-gam[1]}
R[K+1,1]=-2*(gam[1]*y0+f1+gam[6]*x[1] )
R[K+2,c(2,1,K+1)]=c(-f1-gam[1]*y0-gam[6]*x[1],-gam[6]*x[2]-
f1+gam[1]^2*y0+gam[1]*f1+gam[1]*gam[6]*x[1],-gam[1])
for (r in 3:K)
{R[K+r,c(r,1,r-1,K+r-1)]=c(-f1-gam[1]*y0-gam[6]*x[1],-f1-gam[6]*x[r],
gam[1]^2*y0+gam[1]*f1+gam[1]*gam[6]*x[1],-gam[1])}


for (s in 2:(K-1)){
  R[s*K-(s-2)*(s-1)/2+1,
    c(2+(s-1)*K-(s-3)*(s-2)/2,1+(s-1)*K-(s-3)*(s-2)/2,
      s,s-1)]=c(-2*gam[1],gam[1]^2,-2*f1-2*gam[6]*x[s],2*gam[1]*f1+
      2*gam[1]*gam[6]*x[s])


  for (r in (s+1):K){
    R[r-s+s*K-(s-2)*(s-1)/2+1, c(
      r-s+2+(s-1)*K-(s-3)*(s-2)/2,
      r-s+s*K-(s-2)*(s-1)/2,
      r-s+1+(s-1)*K-(s-3)*(s-2)/2,
      r,s,r-1,s-1)]=c(
-gam[1],-gam[1],gam[1]^2,-f1-gam[6]*x[s],-f1-gam[6]*x[r],gam[1]*f1+
        gam[6]*x[s]*gam[1],gam[1]*f1+gam[6]*x[r]*gam[1])
  }
}
```

```
  R[K*K-(K-2)*(K-1)/2+1,
    c(2+(K-1)*K-(K-3)*(K-2)/2,1+(K-1)*K-(K-3)*(K-2)/2,
      K,K-1)]=c(-2*gam[1],gam[1]^2,-2*f1-2*gam[6]*x[K],2*gam[1]*f1+
      2*gam[6]*x[K]*gam[1])

  for (s in 2:(K-1)){
    R[s*K+2-(s-1)*(s-2)/2,s]=-f1-gam[6]*x[s+1]+gam[1]*f1+
    gam[6]*x[s]*gam[1]

  }
  return(R)
}

#This  function is used to calculate  optimal weight matrix of SLS
fw<-function(y,gam,U)
{
  R=Trans(t,y,gam)
  W=t(R) %*% U %*% R
  return(W)
 }


############################################
#Define the output matrix
################################################
```

```
output=array(NA,dim=c(ss,7,7))
dimnames(output)=list(seq(1,ss),c('alpha',
'beta','theta0','theta1','theta2'
,'theta3','theta4'),c('RMLgam','GMMgam','True', 'Empoptsls',
'SLSoptCGLS',
'RMLgam1','GMMgam1'))#the target estimators are GMMgam and Empoptsls
output[,,3]=matrix(1,ss,1)%*%gamm0

set.seed(t*t+n*100*gamm0[1])

############################################################
# This is to generate the fixed x or equivalently (exi) for all
#simulations like kiviet (X is fixed) and using his notation
#####################

#note we tried also to generate  x in a stochastic version, i.e, different
# accross simulation runs and it doesn't make difference

A=array(NA,dim=c(t+1,n,2))# for phi, and psi
B=array(NA,dim=c(t+1,n,2))# for y, and x
eta=rep(NA,n)

for (i in 1:n){
  xi=rnorm(t+1,0,sqrt(sigm2_exi))

  B[1,i,2]=xi[1]/sqrt(1-pro^2)  #the initial value of x
  B[2:(t+1),i,2]=filter(xi[-1],pro,'rec',init=B[1,i,2]) #the rest of x
```

```
  A[1,i,1]=xi[1]*sqrt(varphi_0/sigm2_exi) #the initila value of phi
  A[2,i,1]=A[1,i,1]*p1+xi[2]*sqrt(varphi_0/sigm2_exi)*
  sqrt(1-p1^2) #the in
  itial value of phi
  A[3:(t+1),i,1]=filter(xi[-c(1,2)],c(alpha+pro,-alpha*pro),'rec',init=
  c(A[2,i,1],A[1,i,1])) #the rest of phi
}

for (s in 1:ss){

  for (i in 1:n){

    epsilon=rnorm(t+1,0,sqrt(vareps))

    A[1,i,2]=epsilon[1]/(1-alpha^2)^.5  #the initial value of epsi
A[2:(t+1),i,2]=filter(epsilon[-1],alpha,'rec',init=A[1,i,2])
#the rest of epsi

    eta[i]=rnorm(1,0,vareta^.5)
    B[,i,1]=beta*A[,i,1]+A[,i,2]+eta[i]/(1-alpha)   #the values of y

    }#end of the data generating loop

output[s,1:2,6]=
  nlm(cgls,c(alpha,beta),print.level=1,iterlim=200)$estimate
  #this the RML As in Arelano 2003
```

```
#Computing the  GMM1 (Instruments of Arlano1991)

gmm=apply(rbind(B[,,1],B[,,2]),MARGIN=2,GMM)

tt=(t-1)*t/2+1

Winv=solve(matrix(apply(gmm[-c(1:(3*tt)),],MARGIN=1,sum),tt,tt))

DffX=apply(gmm[1:tt,],MARGIN=1,sum)

DffY=apply(gmm[(tt+1):(2*tt),],MARGIN=1,sum)

DffY_1=apply(gmm[(2*tt+1):(3*(tt)),],MARGIN=1,sum)

R=rbind(DffY_1,DffX)

output[s,1:2,7]=solve(R%*%Winv%*%t(R),R%*%Winv%*%DffY)#this is the GMM1



###################

#Another (more elegent)way to calculate

the GMM1 (Instruments of

#Arlano1991)Stacking the data of n subject in dataframe with

#subject index

################################################

C=data.frame(array(NA,dim=c(n*(t+1),4)))

names(C)=c('subject','time','y','x')#staking the data with index

for (i in 1:n){

  C[(1+(i-1)*(t+1)):(i*(t+1)),]=cbind(i,seq(t+1),B[1:(t+1),i,1],

  B[1:(t+1),i,2])

}

E <- pdata.frame(C, index = c("subject", "time"),
```

```
drop.index = TRUE, row.names = TRUE)


output[s,1:2,2]=pgmm(formula = y ~ lag(y, 1) +
                       x | lag(y,


###########################################
#Another way(more elegent) to comupte the CGLS
#########################

CC=data.frame(array(NA,dim=c(n*(t),7)))
names(CC)=c('subject','time','y','x','y0','y_1','x1')#staking the data
for (i in 1:n){
  CC[(1+(i-1)*(t)):(i*(t)),]=cbind(i,seq(t),B[2:(t+1),i,1],B[2:(t+1),i,2],
  B[1,i,1],B[1:(t),i,1],B[2,i,2])
}

CGLS= lme(y~y_1+y0+x+x1, data=CC,random=~1|subject,control=
list(maxIter=500,returnObject=T) ,method='ML')


output[s,,1]=c(fixef(CGLS)[2],fixef(CGLS)[4],log(CGLS$sigma^2),
fixef(CGLS)[1], fixef(CGLS)[3],
log(coef(CGLS$modelStruct$reStruct,unconstrained=F)*CGLS$sigma^2),
fixef(CGLS)[5])#this the RML
```

```
#compute the residuals to use it latter to compute the third and forth
#moments of the error components appearing in the optimal weight
#matrix W_i
residuals=CC$y-fixef(CGLS)[1]-CC$y_1*fixef(CGLS)[2]-
CC$y0*fixef(CGLS)[3]-CC$x*fixef(CGLS)[4]-CC$x1*fixef(CGLS)[5]

##############################################################
##computing the optimal SLS with Ui evaluated at cgls using
#the estimated third and fourth moments

###########################################################

gam=output[s,c(1,3:6,2,7),1]
#using the estimated third and fourth moments
ETA=tapply(residuals,CC$subject,FUN='mean')
EPS=residuals-rep(ETA,each=t)
ETA=scale(ETA,scale=FALSE)
EPS=scale(EPS,scale=FALSE)
m3eta=mean(ETA^3);m3eps=mean(EPS^3);
m4eta=mean(ETA^4);m4eps=mean(EPS^4)


#define the auxilary matrices to compute the optimal weight  for SLS
###############################################

L <- duplication.matrix( t )
Linv=ginv(L)
a=matrix(1,t,1)
```

```
aa=matrix(1,t^2,1)
A1=matrix(0,t,t^2)
A3=kronecker( diag(1,t),matrix(1,t,t))
A6=vec(diag(1,t))
A4=A6%*%matrix(1,1,t^2)
A2=kronecker(matrix(1,t,t), diag(1,t))
A5=kronecker(t(a),kronecker(diag(1,t),a))
A1[,seq(1,t^2,by=t+1)]<-diag(m3eps,t)
U2=(m3eta*a%*%t(aa)+A1)%*%t(Linv)
A9=commutation.matrix(t,t)


U1=exp(gam[2])*diag(t)+exp(gam[5])*matrix(1,t,t)
A7=vech(U1)
U3<-Linv%*%(
  m4eta*matrix(1,t^2,t^2)+
    exp(gam[2])*exp(gam[5])*(A2+A3+A4+t(A4)+A5+t(A5))+
    exp(2*gam[2])*(diag(1,t^2)+A9+A6%*%t(A6)+
    diag(c(vec(diag((m4eps*exp(-2*    gam[2])-3),t))))))%*%
    t(Linv)-A7%*%t(A7)
    if (eigenvalue<-min(eigen(cbind(rbind(U1,t(U2)),rbind(U2,U3)),
    symmetric=T,
 only.values = T)$value)>0)

{

  U=solve(cbind(rbind(U1,t(U2)),rbind(U2,U3)))
```

```
  W=apply(rbind(B[,,1],B[,,2]),MARGIN=2,FUN=fw,gam,U)
    SLS=nlm(fuu,gamm0,W)
  output[s,,4]=SLS$estimate

}
}#end of simulation runs
save(output,sigm2_s,t,alpha,pro,mu,beta,n, file = 'output.Rdata')
```

Bibliography

Abarin, T. and L. Wang (2009). Second-order least squares estimation of censored regression models. *Journal of Statistical Planning and Inference 139*(2), 125–135. (Cited on page 3.)

Abarin, T. and L. Wang (2012). Instrumental variable approach to covariate measurement error in generalized linear models. *Annals of the Institute of Statistical Mathematics 64*(3), 475–493. (Cited on pages 3 and 4.)

Alvarez, J. and M. Arellano (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica 71*(4), 1121–1159. (Cited on pages 41, 46, 52, 54 and 60.)

Amemiya, T. (1985). *Advanced econometrics*. Harvard university press. (Cited on page 88.)

Arellano, M. (2003). *Panel data econometrics*. OUP Catalogue. (Cited on pages 39, 41, 52 and 72.)

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies 58*(2), 277–297. (Cited on pages 39, 52, 61, 62, 63 and 70.)

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68*(1), 29–51. (Cited on page 39.)

Arellano, M. and B. Honoré (2001). Panel data models: some recent developments. *Handbook of Econometrics 5*, 3229–3296. (Cited on page 39.)

Baltagi, B. (2008). *Econometric analysis of panel data*. Wiley. com. (Cited on page 39.)

Bickel, P. J., C. A. Klaassen, P. J. Bickel, Y. Ritov, J. Klaassen, J. A. Wellner, and Y. Ritov (1993). *Efficient and adaptive estimation for semiparametric models*. Johns Hopkins University Press Baltimore. (Cited on pages 5 and 48.)

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*(1), 115–143. (Cited on pages 39, 40, 41, 46, 54, 60, 64 and 72.)

Bollerslev, T. and J. M. Wooldridge (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric reviews 11*(2), 143–172. (Cited on page 18.)

Chamberlain, G. (1984). Panel data. *Handbook of Econometrics (Z. Griliches and M. Intriligator, eds.)*. (Cited on page 39.)

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics 34*(3), 305–334. (Cited on pages 40, 48 and 79.)

Chamberlain, G. (1992). Comment: sequential moment restrictions in panel data. *Journal of Business & Economic Statistics 10*(1), 20–26. (Cited on page 48.)

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians: An Introduction for Econometricians*. Oxford university press. (Cited on page 86.)

Durairajan, T. (1992). Optimal estimating function for non-orthogonal model. *Journal of statistical planning and inference 33*(3), 381–384. (Cited on page 20.)

Enders, W. (2010). *Applied econometric time series*. John Wiley & Sons. (Cited on pages 9 and 32.)

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society 50*, 987–1007. (Cited on pages 8, 10, 32, 33 and 36.)

Engle, R. F. and G. Gonzalez-Rivera (1991). Semiparametric arch models. *Journal of Business & Economic Statistics 9*(4), 345–359. (Cited on page 8.)

Franses, P. H. and D. Van Dijk (2000). *Non-linear time series models in empirical finance*. Cambridge University Press. (Cited on page 9.)

Godambe, V. P. (1985). The foundations of finite sample estimation in stochastic processes. *Biometrika 72*(2), 419–428. (Cited on page 8.)

Hahn, J. (1997). Efficient estimation of panel data models with sequential moment restrictions. *Journal of Econometrics 79*(1), 1–21. (Cited on page 48.)

Hsiao, C. (2003). *Analysis of panel data*, Volume 34. Cambridge university press. (Cited on page 39.)

Hsiao, C. (2011). dynamic panel data models. *Handbook of Empirical Economics and Finance*, 373–396. (Cited on page 39.)

Ibragimov and Linnik (1971). *Independent and stationary sequences of random variables.* Wolters-Noordhoff. (Cited on page 12.)

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics 40*(2), 633–643. (Cited on pages 83, 88 and 89.)

Kim, M. and Y. Ma (2012). The efficiency of the second-order nonlinear least squares estimator and its extension. *Annals of the Institute of Statistical Mathematics 64*(4), 751–764. (Cited on page 5.)

Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of econometrics 68*(1), 53–78. (Cited on pages v, vi, 60, 62, 63 and 64.)

Li, D. X. and H. Turtle (2000). Semiparametric arch models: an estimating function approach. *Journal of Business & Economic Statistics 18*(2), 174–186. (Cited on pages 8 and 9.)

Li, H. and L. Wang (2012a). Consistent estimation in generalized linear mixed models with measurement error. *Journal of Biometrics & Biostatistics S7:007.* (Cited on page 4.)

Li, H. and L. Wang (2012b). A consistent simulation-based estimator in generalized linear mixed models. *Journal of Statistical Computation and Simulation 82*(8), 1085–1103. (Cited on page 4.)

Liang, Y., A. Thavaneswaran, and B. Abraham (2011). Joint estimation using quadratic estimating function. *Journal of Probability and Statistics 2011*, 14 pages. (Cited on page 8.)

Lindner, A. M. (2009). *Stationarity, mixing, distributional properties and moments of GARCH (p, q)–processes*, pp. 43–69. Handbook of financial time series. Springer. (Cited on page 12.)

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics 4*, 2111–2245. (Cited on page 46.)

Okui, R. (2009). The optimal choice of moments in dynamic panel data models. *Journal of Econometrics 151*(1), 1–16. (Cited on pages 54 and 60.)

Park, B. U., R. C. Sickles, and L. Simar (2007). Semiparametric efficient estimation of dynamic panel data models. *Journal of Econometrics 136*(1), 281–301. (Cited on page 48.)

Rao, C. (1973). *Linear statistical inference and its applications.* New York, NY (EUA). J. Wiley. (Cited on pages 85 and 87.)

Wang, L. (2003). Estimation of nonlinear berkson-type measurement error models. *Statistica Sinica 13*(4), 1201–1210. (Cited on pages 1 and 81.)

Wang, L. (2004). Estimation of nonlinear models with berkson measurement errors. *The Annals of Statistics 32*(6), 2559–2579. (Cited on pages 1 and 81.)

Wang, L. (2007). A unified approach to estimation of nonlinear mixed effects and berkson measurement error models. *Canadian Journal of Statistics 35*(2), 233–248. (Cited on pages 2 and 43.)

Wang, L. and A. Leblanc (2008). Second-order nonlinear least squares estimation. *Annals of the Institute of Statistical Mathematics 60*(4), 883–900. (Cited on pages 3, 5, 89 and 91.)

Weiss, A. A. (1986). Asymptotic theory for arch models: estimation and testing. *Econometric theory 2*, 107–131. (Cited on pages 8 and 18.)

White, H. (1996). *Estimation, inference and specification analysis.* Cambridge university press. (Cited on pages 11, 13 and 17.)

White, H. (2001). *Asymptotic theory for econometricians.* Academic press New York. (Cited on page 86.)

White, H. and I. Domowitz (1984). Nonlinear regression with dependent observations. *Econometrica: Journal of the Econometric Society 52*, 143–161. (Cited on pages 12, 13 and 84.)

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data.* MIT press. (Cited on pages 67 and 70.)